Radboud Repository

Radboud University Nijmegen

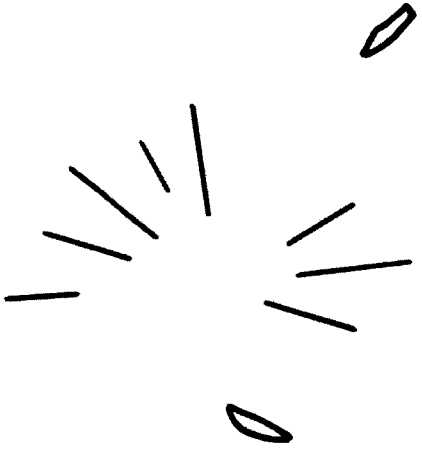
PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.
http://hdl.handle.net/2066/113432

Please be advised that this information was generated on 2016-02-24 and may be subject to change.

# NUCLEIC ACID BINDING PROPERTIES OF THE GENE-5 PROTEIN ENCODED BY BACTERIOPHAGE M13



Henk Bulsink

# NUCLEIC ACID BINDING PROPERTIES OF THE GENE-5 PROTEIN ENCODED BY BACTERIOPHAGE M13

# NUCLEIC ACID BINDING PROPERTIES OF THE GENE-5 PROTEIN ENCODED BY BACTERIOPHAGE M13

Proefschrift
ter verkrijging van de graad van doctor
in de Wiskunde en Natuurwetenschappen
aan de Katholieke Universiteit te Nijmegen,
op gezag van de Rector Magnificus
Prof. Dr. B.M.F. van Iersel
volgens besluit van het College van Decanen
in het openbaar te verdedigen
op vrijdag 20 november 1987
des namiddags te 3.30 uur

door
Hendrik Bulsink
geboren te Dinxperlo.

Druk: Quick Service Drukkerij Enschede

Promotor: Prof. Dr. C.W. Hilbers

Co-referent: Dr. B.J.M. Harmsen

I.S.B.N. 90-900-1428-4

The investigations reported in this thesis were supported by the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Organisation for the Advancement of Pure Research (ZWO). The award of an EMBO fellowship by ZWO is gratefully acknowledged.

opgedragen aan mijn ouders

## CONTENTS


1. INTRODUCTION . . . . . .11
   1.1 Bacteriophage M13 . . . . . .11
      1.1a The structure of bacteriophage M13 . . . . . .11
      1.1b Phage replication . . . . . .13
      1.1c Related bacteriophages . . . . . .16
   1.2 The gene-5 protein . . . . . .16
   1.3 The gene-5 protein DNA complex . . . . . .20
      1.3a Structural characteristics of the gene-5 protein - polynucleotide complex . . . . . .20
      1.3b Structural characteristics of the gene-5 protein - oligonucleotide complex . . . . . .23
      1.3c The binding stoichiometry . . . . . .24
      1.3d The binding cooperativity . . . . . .26
      1.3e Salt dependency of the binding . . . . . .27
      1.3f Association kinetics . . . . . .27
      1.3g Protein residues involved in binding . . . . . .29
   1.4 Outline of this thesis . . . . . .32

2. THEORETICAL ASPECTS . . . . . .35
   2.1 Binding theories . . . . . .35
      2.1a Single mode binding to polymers . . . . . .35
      2.1b Single mode binding to oligomers . . . . . .44
      2.1c Multiple mode binding to polymers . . . . . .47
   2.2 Kinetic theories . . . . . .49
      2.2a The non-redistributing model . . . . . .51
      2.2b The immediately-redistributing model . . . . . .54

3. SPECIFICITY OF THE BINDING OF BACTERIOPHAGE M13 ENCODED GENE-5 PROTEIN TO DNA AND RNA STUDIED BY MEANS OF FLUORESCENCE TITRATIONS . . . . . .57
   3.1 Abstract . . . . . .57
   3.2 Introduction . . . . . .57
   3.3 Materials and methods . . . . . .59
      3.3a Materials . . . . . .59
      3.3b Fluorescence titrations . . . . . .59

3.3c Evaluation of titration curves . . . . . .60
3.4 Results . . . . . .63
3.4a Titration experiments . . . . . .63
3.4b Derivation of the cooperativity parameter from salt titrations .69
3.4c Temperature dependency of the complex formation . . . . . .72
3.5 Discussion . . . . . .73
3.5a Binding characteristics . . . . . .73
3.5b Specificity of the binding. . . . . .74
3.5c Comparison with the gene-32 protein . . . . . .76
3.5d Possible implications for the in vivo behaviour of the gene-5 protein . . . . . .77

4. DIFFERENT DNA BINDING MODES AND COOPERATIVITIES FOR BACTERIOPHAGE M13 GENE-5 PROTEIN REVEALED BY MEANS OF FLUORESCENCE DEPOLARISATION STUDIES .85
4.1 Abstract . . . . . .85
4.2 Introduction . . . . . .86
4.3 Materials and methods. . . . . .87
4.3a Materials . . . . . .87
4.3b Fluorescence measurements . . . . . .88
4.4 Results . . . . . .89
4.4a Fluorescence lifetime measurements . . . . . .89
4.4b Fluorescence depolarisation measurements. . . . . .90
4.4c Titration experiments . . . . . .92
4.5 Discussion . . . . . .94
4.6 Appendix . . . . . 100

5. THE KINETICS OF THE DISSOCIATION OF GENE-5 PROTEIN - POLYNUCLEOTIDE COMPLEXES UPON ADDITION OF SALT . . . . . 103
5.1 Abstract . . . . . 103
5.2 Introduction . . . . . 104
5.3 Materials and methods. . . . . 106
5.3a Materials . . . . . 106
5.3b Stopped flow experiments. . . . . 106
5.3c $^{1}H$ NMR spectra . . . . . 109
5.4 Results . . . . . 109
5.4a Stopped flow measurements of the irreversible dissociation of gene-5 protein - poly(dA) and gene-5 protein - M13 DNA

complexes . . . . . 109
5.4b Effect of binding of gene-5 protein on the $^{1}H$ NMR spectra of polynucleotides . . . . . 113
5.5 Discussion . . . . . 114

6. FURTHER CHARACTERIZATION OF THE DIFFERENT BINDING MODES FOR POLY- AND OLIGODEOXYNUCLEIC ACIDS . . . . . 125
6.1 Abstract . . . . . 125
6.2 Introduction . . . . . 126
6.3 Materials and methods. . . . . 127
6.3a Materials . . . . . 127
6.3b Fluorescence binding experiments . . . . . 128
6.3c Fluorescence depolarisation measurements . . . . . 128
6.3d Stopped flow irreversible dissociation experiments . . . . . 129
6.4 Results . . . . . 129
6.4a Binding parameters from fluorescence titrations for the binding to polynucleotides . . . . . 129
6.4b Binding parameters from fluorescence titrations for the binding to oligonucleotides . . . . . 132
6.4c Fluorescence depolarisation experiments . . . . . 137
6.4d Stopped flow dissociation kinetics . . . . . 141
6.5 Discussion . . . . . 144
6.5a Dual mode binding of gene-5 protein to polynucleotides. . . . . 144
6.5b Binding of gene-5 protein to oligonucleotides . . . . . 146
6.5c The n = 3 modes do not represent the same binding in oligo- and polynucleotides . . . . . 150
6.5d Influence of the nucleotide chain length on the apparent binding constant and the effective dissociation rate of oligo(dT) - gene-5 protein complexes . . . . . 154
6.6 Appendix . . . . . 155

7. A TWO DIMENSION NMR STUDY OF THE BACTERIOPHAGE M13 GENE-5 PROTEIN HEXADEOXYADENYLIC ACID COMPLEX . . . . . 161
7.1 Abstract . . . . . 161
7.2 Introduction . . . . . 161
7.3 Materials and methods . . . . . 162
7.3a Materials . . . . . 162

7.3b NMR spectra . . . . . 163
7.4 Results . . . . . 164
7.4a The $d(A)_6$ proton spectrum . . . . . 164
7.4b Complex formation between $d(A)_6$ and the gene-5 protein . . . . 168
7.5 Discussion . . . . . 174
7.6 Appendix . . . . . 179
7.6a Method to calculate peak intensities in two-dimensional exchange spectra. . . . . 179
7.6b Cross peak intensities in two-dimensional exchange spectra of a model system . . . . . 182

REFERENCES . . . . . 187
SUMMARY . . . . . 197
SAMENVATTING . . . . . 200
DANKWOORD . . . . . 203
CURRICULUM VITAE . . . . . 204

Chapter 3:
Bulsink, H., Harmsen, B.J.M. and Hilbers, C.W.,
J. Biomol. Struct. Dyn. (1985) 3, 227-247

Chapter 4:
Bulsink, H., Wijnaendts van Resandt, R.W., Harmsen, B.J.M. and Hilbers, C.W.,
Eur. J. Biochem. (1986) 157, 329-334

Chapters 5 and 6:
submitted for publication in J. Biol. Chem.

Chapter 7:
submitted for publication in Nucl. Acids Res.

# CHAPTER 1

# INTRODUCTION

The gene-5 protein of bacteriophage M13, which forms the subject of this thesis, is an example of the class of non-sitespecific single stranded DNA binding proteins. For several reasons it is attractive to study the nucleic acid binding properties of this protein. It is a small molecule compared to most nucleic acid binding proteins and therefore can serve as a relative simple model system. Furthermore, a significant amount of structural and functional information is now available for the protein (as well for the bacteriophage M13) so that rather penetrating questions can be asked with respect to its structure / function relationships.

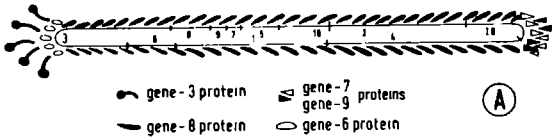
In the next sections a review will be given of literature which deals with bacteriophage M13 and its gene-5 protein. Finally an outline of this thesis will be given.

## 1.1 BACTERIOPHAGE M13

From an incubation of waste water samples with several types of Escherichia coli bacteria, Hofschneider discovered in 1963 some new types of bacteriophages, among which was bacteriophage M(unich) 13 [1]. Since then this bacteriophage has been extensively studied by means of genetical, biochemical and physical techniques. Nowadays its DNA is widely used as a vector in DNA recombinant experiments.

### 1.1a The structure of bacteriophage M13

A schematic drawing of the bacteriophage, as derived from biochemical and physical studies [2], is given in Fig. 1.1A. M13 has a threadlike form (it is



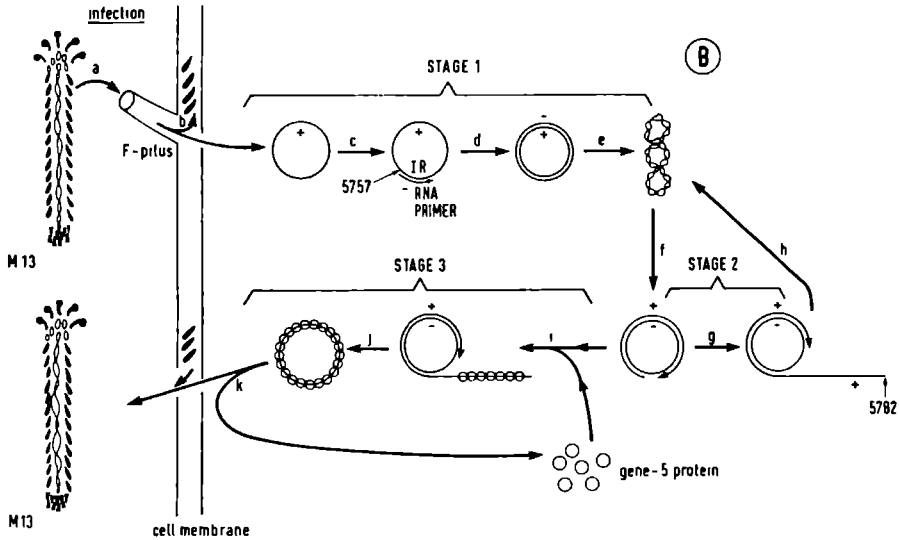




Figure 1.1. A. Schematic representation of the structure of the M13 phage. The relative positions of the genes are indicated. Gene 10 is part of gene 2. The intergenic region (IR) is positioned between gene 2 and gene 4. The figure only shows the relative location of protein molecules and does not imply any interactions among proteins or between proteins and the DNA.
B. Scheme for M13 DNA infection (reaction a), the three stages of phage replication (see text, reactions c to j) and phage morphogenesis (reaction k). The plus and minus sign denote the viral and complementary strands. The positions 5757 and 5782 on the genome are the starting points for respectively complementary and viral strand synthesis.

9850 Å long and only 60 Å wide) and is classified as one of the filamentous bacteriophages. The phage genome consists of a circular single-stranded DNA, 6407 nucleotides long, of which the sequence has been determined [3]. The genomic organisation is also indicated in Fig. 1.1A. The genome consists of 10 genes (after which the protein products are named) and a non-coding intergenic region. The DNA is predominantly encapsulated by approximately 2700 molecules of the gene-8 protein (also called the major coat protein or B-protein). Thus the number of nucleotides corresponding with one coat protein amounts to about 2.3. The circular single stranded DNA in the center is elongated along the axial length of the bacteriophage and occupies a core of less than 25 Å in diameter [4,5]. Apart from the major coat protein four types of minor coat proteins are located at the tips of the phage (see Fig. 1.1A) [2,6,7]. One extremity is formed by about a total of ten molecules of the gene-7 and gene-9 proteins (also called C-proteins). At this end of the virus the mentioned intergenic region of the phage genome is found. At the other tip about five molecules of the gene-3 protein (also designated A-protein) and about five molecules of the gene-6 protein (D-protein) are situated (see Fig. 1 1A)

### 1.1b Phage replication

Bacteriophage M13 infects E. coli bacteria with F-type pili, which are hairlike structures at the outside of the E. coli cell. The gene-3 protein at one end of the phage, is involved in adsorption of the phage to the E. coli cell [8,9] (therefore it is also called the pilot protein or attachment protein). During infection the virus penetrates the host bacterium and, when passing the membrane, the major coat protein is stripped off and deposited into the cell membrane for later use [10]. After penetration the replication process is started. This process consists of three stages (see Fig. 1.1B): a) the conversion of viral DNA into a circular duplex form, called the replicative form (RF), b) the RF multiplication and c) the single strand synthesis [11]. Below we will consider each stage separately.

In the first stage of the replication process at least six E. coli encoded enzymes are needed [12] Before the actual start of the process the single-stranded DNA is probably masked by the E. coli SSB protein as a result of

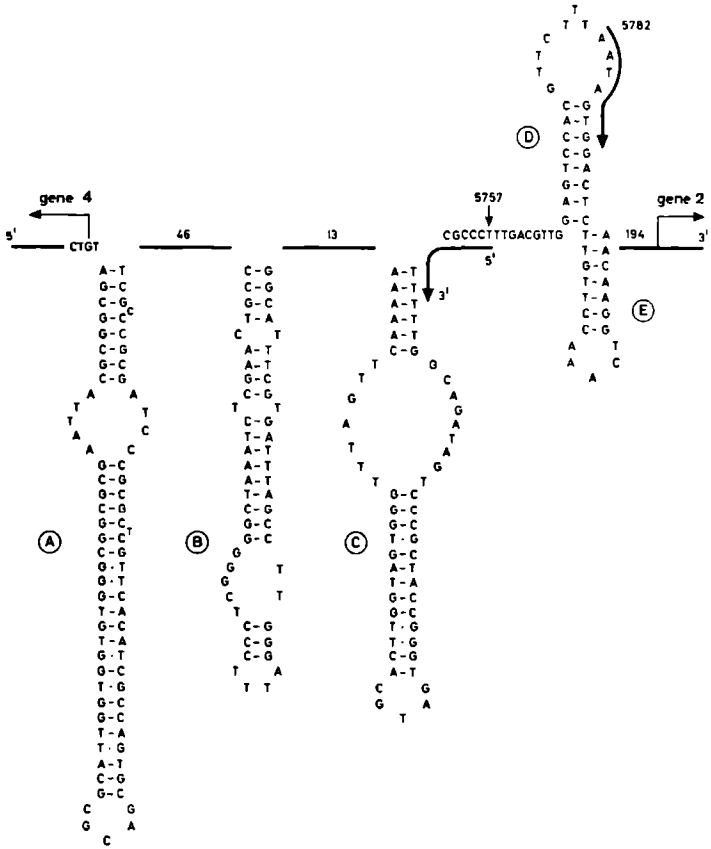



Figure 1.2. The intergenic region of the M13 genome with its proposed secondary structure (hairpins A to E). The positions 5757 and 5782 are the starting points for respectively complementary and viral strand synthesis, denoted by bold arrows. Single stranded sequences containing no secondary structure are represented by straight lines with indication of the number of nucleotides composing these sequences.

which the phage DNA is straightened and most hairpins are melted, except for the double helical region (hairpin C, see Fig. 1.2) which serves as a promotor region for RNA polymerase [13]. The complementary strand synthesis is initiated by RNA polymerase; the molecule attaches specifically to the unmasked region of the DNA and synthesizes a short complementary RNA chain. This RNA is utilized as a primer for DNA synthesis by DNA polymerase III and DNA copolymerase III. After replication of the complete DNA the primer RNA is removed by DNA polymerase I. DNA ligase closes the newly synthesized strand to form a circular double stranded DNA molecule. Subsequently, DNA gyrase converts the double stranded DNA into a superhelical structure, the parental replicative form DNA (RF-I).

In the second stage of the replication process multiplication of the RF duplex takes place. It is generally assumed that this step occurs via the so-called "rolling circle" mechanism. In the process the gene-2 protein nicks the parental strand in the RF duplex at a unique site in the intergenic region (see caption of Fig. 1.2). Subsequently, replication takes place at the 3' nicked end, while the viral strand is displaced from the 5' end, then is closed and serves as a template for complementary strand synthesis. The duplex is closed by the gene-2 protein. The products of one round of plus-strand synthesis are a circular single strand and a double stranded circle [14]. In vitro studies have shown that the duplex replication process is markedly stimulated by the addition of purified E. coli single strand binding protein [15], from which the conclusion can be drawn that this protein is involved in both the enzymatic melting of the duplex DNA and in converting the unwound DNA into a suitable template for DNA polymerase III.

At about 15 to 20 minutes after infection the transition from RF replication to single strand synthesis takes place [16]. At this point of time the RF pool has grown to 100 to 200 copies. In this last stage of the replication process viral strands are synthesized. The presence of the gene-5 protein is required in this process: it binds to the displaced viral strands produced in the "rolling circle" process and thereby prevents complementary strand synthesis [17,18]. The transition is initiated after a sufficient amount of gene-5 protein has become available and the moment at which it starts is not determined by the size of the RF pool [19]. The necessary amount of gene-5 protein is estimated to be about $1.2 \times 10^5$ molecules per infected cell [18].

Finally, the viral DNA - gene-5 protein complex migrates into the cell membrane. Upon passing the membrane the phage coat protein, which has accumulated in the membrane, replaces the gene-5 protein. The gene-5 protein molecules are recycled into the cell, where they can associate with new single stranded DNA molecules [20]. For a more extensive description of the subject of this section the reader is referred to three reviews [14,21,22].

### 1.1c Related bacteriophages

The different groups of filamentous bacteriophages can be classified on the basis of their specificity for bacterial cells [23]. The Ff group, which is specific for E. coli cells carrying the F-type sex pilus, includes amongst others the bacteriophages fd, f1, and M13 [24]. On comparing the sequences of the phage DNA molecules, it appears that M13 and f1 are more closely related to each other than M13 and fd [25-27]. The homology between the M13 DNA and the f1 DNA amounts to 99.2 %, while it is 97 % between the M13 DNA and the fd DNA. No differences in the sequence are found in that part of the intergenic region in which the starting points of complementary strand and viral strand synthesis are situated (see Fig. 1.2). The genome of bacteriophage fd is one nucleotide longer than the genomes of bacteriophage M13 and f1.

## 1.2 THE GENE-5 PROTEIN

The gene-5 protein was first detected in crude extracts of E. coli cells infected by M13 phages [28]. Later on the protein could be isolated from infected E. coli cells using chromatography on single stranded DNA cellulose [17,29]. The protein consists of 87 amino acids [30,31], the gene-5 proteins encoded by the bacteriophages f1, fd and M13 have the same amino acid sequence [27,30,31], which is as follows:

Met-Ile-Lys-Val-Glu-Ile-Lys-Pro-Ser-Gln(10)-
Ala-Gln-Phe-Thr-Thr-Arg-Ser-Gly-Val-Ser(20)-
Arg-Gln-Gly-Lys-Pro-Tyr-Ser-Leu-Asn-Glu(30)-
Gln-Leu-Cys-Tyr-Val-Asp-Leu-Gly-Asn-Glu(40)-
Tyr-Pro-Val-Leu-Val-Lys-Ile-Thr-Leu-Asp(50)-

Glu-Gly-Gln-Pro-Ala-Tyr-Ala-Pro-Gly-Leu(60)-
Tyr-Thr-Val-His-Leu-Ser-Ser-Phe-Lys-Val(70)-
Gly-Gln-Phe-Gly-Ser-Leu-Met-Ile-Asp-Arg(80)-
Leu-Arg-Leu-Val-Pro-Ala-Lys(86)

This sequence accounts for a molecular weight of 9688 g per mol. The isoelectric point is at least 8.0 [17]. In solution the protein is present as a dimer [18,32-35], which is stable between 0 to 0.5 M KCl and pH 5.0 to 11.0 [36]. From equilibrium sedimentation curves a dimer stability constant of 8.0 x $10^5$ $M^{-1}$ can be calculated [35]. This means that, at an estimated concentration of $10^{-4}$ M gene-5 protein in the E. coli cell [37], about 92 % of the protein exists in the dimer form.

In solution, the gene-5 protein shows a tendency to form aggregates. This tendency increases with the temperature. For example, protein solutions cannot be kept for more than 15 minutes at about 35 °C without denaturation [own observations]. Protein solutions are also sensitive to turbulences caused by shaking or stirring [33]. Protein aggregation leads to appreciable light scattering since the decrease of light transmission due to scatter is stronger at lower wavelengths. The ratio of the UV absorption at 280 nm to the absorption at 260 nm can be used as an indication of the degree of aggregation. In practice the upper value of this ratio is 2.1 which is taken to be representative for native protein. In contrast, at values below 1.8 the gene-5 protein has lost its affinity for DNA.

Gene-5 protein crystals have been subjected to X-ray diffraction studies [38-41]. These investigations have led to a 2.3 Å resolution of the gene-5 protein structure. The molecular dimensions of the dimer have been approximated to be 55 Å x 45 Å x 36 Å. In Fig. 1.3 the gene-5 protein structure, as obtained from these crystallographic studies, is visualized. The protein monomers are composed of a three-stranded β-sheet (the three parallel white strands in Figs. 1.3B and 1.3C), which projects outwards from a common hydrophobic core (dark strands in Figs. 1.3B and 1.3C). The latter connects the two dimer halves in a compact way symmetrically about a 2-fold axis.

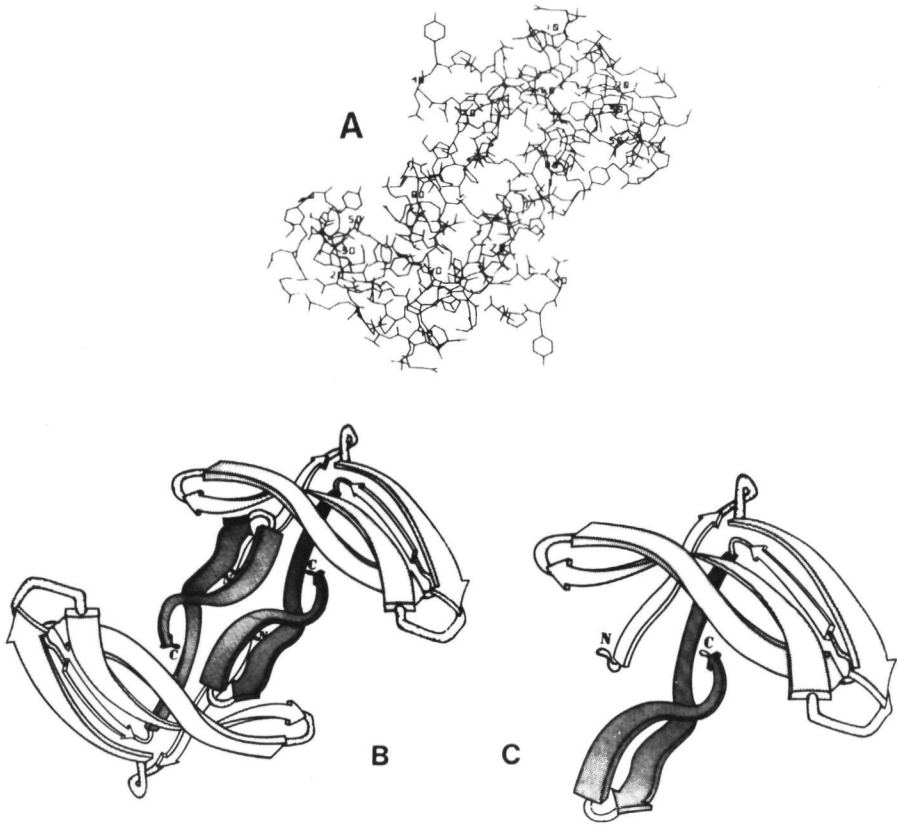



Fig. 1.3. Three dimensional structure of gene-5 protein as obtained from crystallographic studies [41]. In A a drawing of all non-hydrogen atoms of the gene-5 protein dimer is given. A stylized drawing of the polypeptide backbone of the dimer (B) and one protein monomer (C) shows the three-stranded β-loop structure, and the dimer-dimer interaction region, which is formed by the C-terminal part. The numbers refer to amino acid residues; C and N refer to the C and N terminal end of the protein. Reproduced from Brayer and McPherson [41] by copyright permission of the authors and the Journal of Molecular Biology.

The conclusion that the gene-5 protein is needed for the conversion of double stranded RF synthesis to single stranded viral DNA formation [42] is consistent with the observation of a decrease of the melting temperature of double stranded DNA in the presence of gene-5 protein. E.g. upon addition of gene-5 protein the melting temperature of several double stranded DNAs is lowered by about 40 $^{o}$C [17,18,43,44], clearly indicating double helix destabilization. Thus the gene-5 protein has a preference for binding to single stranded DNA. This was also demonstrated in experiments in which gene-5 protein was added to a solution of poly(dA).poly(dT) [45]. In these experiments a quenching of the protein fluorescence, which denotes binding to the DNA, was only observed at sample temperatures above 20 $^{o}$C where the double stranded DNA is melted.

Apart from the double helix destabilizing properties the gene-5 protein also offers protection against digestion. Experiments, in which fd DNA was digested by DNAse I [46] or exonuclease I and exonuclease III [18] showed that digestion was prevented in the presence of gene-5 protein. Thus, during single stranded DNA replication the protein may protect the viral DNA against nucleolytic attack. Protection is not due to inactivation of the nuclease by the gene-5 protein. At higher salt concentrations, at which gene-5 protein does not bind to DNA, DNAse I is still able to hydrolyse DNA in the presence of the gene-5 protein [own observations].

In addition to the gene-5 protein activities discussed above there exists experimental evidence that the gene-5 protein is involved in the translational repression of the M13 encoded gene-2 protein. Addition of gene-5 protein to an in vitro protein-synthesizing system inhibits the synthesis of gene-2 protein, whereas cells infected with a mutant carrying a mutation in gene-5 overproduce gene-2 proteins, which results in a decontrolled phage DNA synthesis. These in vitro and in vivo experiments of Model et. al. [47] are in agreement with in vivo experiments performed by Yen and Webster [48], in which an overproduction of gene-2 protein is observed in the absence of functional gene-5 protein. The same translational repression is found for the gene-10 protein; gene-10 is a part of gene-2 and the translation from gene-10 is in-phase with that of gene-2. The gene-10 protein probably plays a role in the viral strand synthesis [49].

## 1.3 THE GENE-5 PROTEIN DNA COMPLEX

### 1.3a Structural characteristics of the gene-5 protein - polynucleotide complex

So far structural information regarding the gene-5 protein - viral DNA complex has mainly been obtained from electron microscopy [17,20,50,51] and neutron scattering data [50,52]. From these studies and additional information of other studies, i.e. sedimentation equilibrium measurements of the complex by Cavalieri et al. [34] but also from viscosity measurements of the complex by Hollwitt and Krasna [44] and from an X-ray diffraction study of the gene-5 protein and its complex with oligonucleotides by McPherson et al. [39], it has been suggested that each half of a protein dimer binds to a different DNA strand. As a result the two protein-covered strands form a fiber coil (see Fig. 1.4). However, a structure in which both dimer halves bind to the same DNA strand keeping the two strands together by dimer-dimer interaction cannot be excluded. The rod-like complex contains a small amount of branches (found for reconstituted complexes at low ionic strength) as shown in the electron microphotographs of Alberts et al. [17] and of Torbet et al. [50]. In Table 1.1 structural features of the gene-5 protein viral DNA complex are given, as obtained from the literature. Some of the parameters are defined in Fig. 1.4 (left drawing), while the cross sectional dimensions, adapted from the detailed neutron scattering data of Gray et al. [52], are shown in the right hand drawing of Fig. 1.4.

There is global agreement between the various results in Table 1.1 except for the results from the electron micrograph experiments of Pratt et al. [20] which are somewhat deviating, probably due to the use of an ethanol dehydration step, which could have altered the protein structure [51].

From a comparison of the contour dimensions of the gene-5 protein - viral DNA complex (Table 1.1) and the bacteriophage dimension (9850 Å long and 60 Å wide) it can be conluded that there exists a close relationship in overall length between the phage and the gene-5 protein - viral DNA complex This indicates that the gene-5 protein holds the DNA in a conformation in which it can easily be displaced by the coat protein during phage assembly. If we compare the dimensions of the bacteriophage and of the gene-5 protein complex,

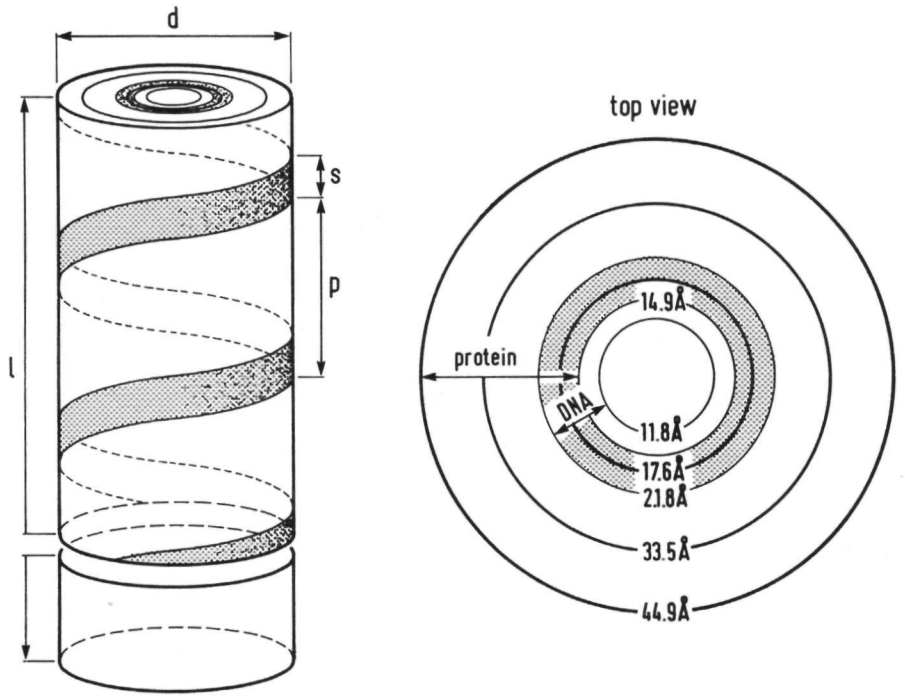



Fig. 1.4. Structure of the gene-5 protein complex helix, based on neutron scattering data of Gray et al. [52]. At the left a view perpendicular to the helical axis of the complex is shown. The parameters ℓ, d s and p (see also Table 1.I) characterize the dimensions of the helix. The stippled region projected on the outside of the cylinder represents the outer dimensions of the complex. On the right a presentation of the complex, viewed from the top, is given. The protein is situated between 11.8 and 44.9 Å; this region overlaps partially the region 11.8 Å to 21.8 Å, in which the DNA is placed. The radii of gyration at infinite neutron scattering contrast are 33.45 Å for the protein and 17.6 Å for the DNA.

Table 1 I.
Structural features of the gene-5 protein - viral DNA complex
The parameters ℓ, d, s, p, $R_p$ and $R_D$ are given in Å.
M denotes the number of nucleotides per strand (M = 6409).

| | | in vivo | | | in vitro | | |
|---|---|---|---|---|---|---|---|
| Reference | | [20] | [50] | [51] | [17] | [50] | [52] |
| contour length | (ℓ) | 11000 | 8800±20 | 8800 | 7300 | 9100±30 | 9410±480<br>8400±350[2] |
| contour diameter | (d) | 160 | 100±10[3]<br>≤100[4] | 95-105 | 100 | 100±10[3]<br>≤100[4] | 89.8 |
| fiber strand thickness | (s) | 40 | -- | -- | 45-50 | -- | -- |
| helical pitch | (p) | 160 | 91 | 91±1[5]<br>90±30[6]<br>100±10[7] | 65-70 | 92 | -- |
| protein radius | ($R_p$) | -- | 34 5±1 | 35 | -- | 34 5±1 | 33 5±0 4 |
| DNA radius | ($R_D$) | -- | ≤21 4 | -- | -- | ≤21.4 | 17 6±3 |
| number of helical turns | ($N_t$) | 69[8] | 96-98 | 97±1 | 100-120 | 97-101 | -- |
| number of nucleotides per turn[1] | ($N_D$) | 46 | 23 | 33 | 30±3[9] | 32±1[9] | -- |
| nucleotide axial translation[10] | (t) | 3.4 | 2.7 | 2.7 | 2.3 | 2.8 | 2.9<br>2 6[2] |

[1] Number of nucleotides per turn for each strand Remember that there are two strands per turn, [2] Deuterated DNA, [3] From electron microscopy, [4] From neutron diffraction; [5] Obtained from $\ell/N_t$, [6] Obtained from electron micrographs. The large range of different values indicates a flexible helix, [7] From measurements of selected straight regular segments in electron micrographs, [8] Obtained from ℓ/p, [9] Obtained from $M/(2N_t)$, [10] Obtained from 2 ℓ/M.

then we see that the outer diameter of the phage (60 Å) is about 2/3 of the complex diameter (see Table 1.1). This is a consequence of the smaller coat protein cylinder, which has an inner radius of 20 Å and an outer radius of 30 Å [5,53,54]. The DNA is located within a central core with a radius of 12.5 Å. This means that in the phage the DNA has a volume which is about half that in the gene-5 protein - viral DNA complex (see Table 1.1 and Fig. 1.4). So the displacement of the gene-5 protein by the coat protein during phage

assembly involves a change in the outer contour dimension of the complex, while the overall length and axial translation hardly change.

Values are reported for the nucleotide axial translations of the virus (2.8 Å [55]) and the gene-5 protein - viral DNA complex (Table 1.1) which are approximately equal for both and lie between the values for A type DNA (2.59 Å) and B type DNA (3.38 Å) [56]. The number of nucleotides per turn is 11 for A-DNA, 10 for B-DNA and about three times higher for the gene-5 protein - DNA complex (see Table 1.1). Taking all this together, it means that also the helical pitch of the gene-5 protein complex must be about three times as large as the helical pitch of A-DNA (28 5 Å) and B-DNA (33.8Å) [56]. Combined with the large range of values observed for the helical pitch from electron micrographs (Table 1.1), this indicates that the gene-5 protein complex must have a strongly open and solvated structure, in which however the DNA is sufficiently protected against nuclease digestion (see above).

The value of the outer DNA radius of natural DNA is about 20 Å [56]. The DNA is situated in the gene-5 protein - viral DNA complex in a shell with an outer radius of 21.8 Å, which is only slightly larger than the value of the outer DNA radius of natural DNA.

### 1.3b Structural characteristics of the gene-5 protein - oligonucleotide complex

By means of co-crystallization and diffusion [39,40,57,58] McPherson et al. tried to form complexes of gene-5 protein with oligonucleotides consisting of 2, 3, 4, 6 or 8 residues. From the unit cell dimensions and a density, which is supposed to be equal to the density in the gene-5 protein crystals, the number of gene-5 protein dimers was found to be 6 per asymmetric unit. Because of the larger size of this asymmetric unit, the atomic cordinates could not be derived from the X-ray diffraction patterns. A model for the gene-5 protein - oligonucleotide complex was proposed, consisting of a circle of six gene-5 protein dimers arranged side by side (disk model). The DNA binding groove is believed to be situated at the outiside of the circle [40]. The difference to the gene-5 protein - polynucleotide complex, which forms a helix (see above) is clear. A conversion from the disk to the helical structure at

longer nucleotide chains has been proposed, in analogy to the structural conversion occurring in Tobacco Mosaic Virus [40]. However, the DNA will then be situated at the outside of the helical complex, which is not supported by the experiments of Gray et al. [52] (see Table 1.1 and right hand of Fig. 1.4).

Electron density maps at a resolution of 5.0 Å [40] or 2.8 Å [58] were obtained for putative complexes prepared with the diffusion technique. The major electron density appeared at the ends of the proposed DNA binding clefts (see Fig. 1.5), i.e. at residues 26-34.

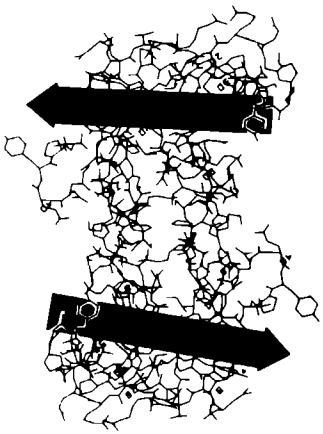


Fig 1 5 Drawing of the gene-5 protein dimer in which the supposed orientations of the two DNA binding channels have been marked with black arrows [5 ] Reproduced from Brayer and McPherson [41] by copyright permission of the authors and the Journal of Molecular Biology

### 1.3c The binding stoichiometry

One of the binding parameters characterizing the complex formation between DNA and the gene-5 protein is the number (n) of nucleotides covered by one protein molecule. Most studies [17,29,33-35,37,43-45,59-64] reveal a binding stoichiometry of four nucleotides per protein monomer for the binding to polynucleotides.

However, also deviations from this value have been reported. Some of these arise from the fact that at the time of the investigation insufficient information was available with respect to the mass and the composition of the protein. For example, in the sucrose-gradient sedimentation experiments, performed by Oey and Knippers [18], the fraction of gene-5 protein bound to fd DNA was measured via tritium labeling of the protein and the optical density of the DNA. It was found in vitro that the total mass of bound gene-5 protein was 7 to 8 times higher than the total mass of DNA, resulting in an estimation of n = 4.5. Since then more precise values for the molecular masses (9688 g $mol^{-1}$ for gene-5 protein [30], 2.12 x $10^6$ g $mol^{-1}$ for fd DNA [26]) and the number of nucleotides per fd DNA (6408 [26]) became available. Recalculation of n with these new data gives values between 3.7 and 4.2, consistent with the other mentioned experiments.

There is another reason why deviations of the general value n = 4 are reported in the literature. As an illustration the proceeding of Pretorius et al. [33] is reported. An in vivo gene-5 protein - fd DNA complex was isolated and the complex was dissociated by increasing the pH to 12.6. Subsequently, the concentrations of DNA and gene-5 protein were determined from the UV absorbance, which produced an average value of n = 4.7. However, in this case the DNA was only partly saturated with protein as followed from a calculation of the degree of binding for the conditions mentioned on the basis of experimentally obtained binding constants [37]. Therefore the actual value of n must have been lower than 4.7. More recent data [50,51,64] show no substantial difference in binding stoichiometry between in vivo and in vitro complexes.

Both sources just mentioned for a deviating value of the stoichiometry play a role in the study of Pratt et al. [20]. Not exact numbers have been taken for the amount of Leu residues per gene-5 protein monomer (compare [30,31]), the number of nucleotides per M13 DNA and the number of thymidines per M13 DNA (compare [3]), upon which numbers the calculation of n was based in this investigation. Moreover the in vivo complex was probably not completely saturated too, resulting in too high values of n.

Taking this all together the final conclusion must be that the normal polynucleotide binding mode is characterized by a binding stoichiometry of four nucleotides per protein monomer. This means that the value of n = 5, used by Brayer and McPherson [58,65,66] in their molecular modelling studies of the M13 DNA gene-5 protein complex, is too high.

In contrast to the n-values exceeding 4, in certain cases a binding stoichiometry of n = 3 has been found. This latter value was obtained in proton NMR studies of the binding of gene-5 protein to short stretches of deoxyadenylic acid (8, 12 and 16 residues in length) [67] and in a CD study of the binding to $d(A)_7$ [68]. A transition from the n = 3 binding mode for oligonucleotides to the n = 4 binding mode for polynucleotides, induced by variation of the nucleotide / protein ratio was observed at a nucleotide chain length of 25 to 30 residues [69]. A recent study of the gene-5 protein polynucleotide binding, based upon circular dichroism measurements of the nucleic acid, showed that binding stoichiometries of n = 3 can also be observed for polynucleotide complexes at very low ionic strength [70].

### 1.3d The binding cooperativity

The occurrence of "humped" Scatchardplots [34,37] and sigmoidal binding isotherms [32,34,37,45] indicate that the gene-5 protein binding to polynucleotides is (positive) cooperative, which means that relative to binding at an isolated site protein molecules bind stronger next to already bound molecules. Further evidence for cooperative binding is provided by cross linking experiments [32,34]. Binding cooperativity can be quantified by the introduction of the so-called cooperativity parameter, $\omega$, which is defined as the increase of the binding constant upon binding of a protein molecule adjacent to an already bound molecule compared to isolated site binding. Values of $\omega$ were obtained from fluorescence binding experiments (see also Chapter 2) [71]. They are estimated to be about 500 [35,37,61,72]. A strong cooperative effect, as observed for the polynucleotide binding, has not been found for the oligonucleotide binding. An estimate of $\omega$ = 5 has been reported from proton NMR titrations of gene-5 protein with oligodeoxyadenylic acids [67].

### 1.3e Salt dependency of the binding

During the isolation procedure of gene-5 protein from infected cells use is made of the binding property that addition of salt may dissociate gene-5 protein - DNA complexes. This means that at higher salt concentrations the binding strength is strongly decreased. Studies of the salt dependence of the binding constant have been limited so far. Binding constants for "isolated" binding to poly(dA), M13 DNA and poly(dT) were determined at about 0.2 M NaCl and appeared to be roughly $10^3$ $M^{-1}$, $10^4$ $M^{-1}$ and > $10^5$ $M^{-1}$ respectively [37]. The salt dependence of the complex formation with poly(dA) has been studied in more detail [37]. These results were found to conform to the theory of Record et al. [73,74] In a doubly logarithmic plot a linear relationship has been observed between the binding constant and the ionic strength. From this plot it was concluded that in the binding process the protein loses one anion while three cations are released from the poly(dA) [37]. Binding of the protein to octa-deoxyadenylic acid is also salt dependent [67], but not as pronounced as the binding to polynucleotides. The binding to oligonucleotides thus differs not only from polynucleotide binding in binding stoichiometry and cooperativity, but also in the salt dependence of the binding constant. As a conclusion it is stated that electrostatic interactions form an important contribution to the stability of gene-5 protein - nucleotide complexes.

### 1.3f Association kinetics

Apart from equilibrium binding studies the complex formation of the gene-5 protein with DNA has also been investigated by means of kinetic studies. These experiments have been performed for the formation of the gene-5 protein poly(dT), poly(rA) and $d(T)_4$ complexes [35] and for the gene-5 protein fd DNA complex [59]. The association rate constant for the isolated binding of gene-5 protein is reported to be 1.6 x $10^{10}$ $M^{-1}$ (fd DNA) $s^{-1}$ and 8.1 x $10^9$ $M^{-1}$ (poly(dT)) $s^{-1}$. These values are in the range expected for a diffusion-controlled association. It is stated that the rate of cooperative association cannot exceed this upper limit; therefore the association rates for isolated site and cooperative binding must be equal. This means that the coop-

erativity only plays a role in the kinetics of the dissociation process; $\omega$ is said to be dissociation-controlled [35,59].

To obtain complete saturation of the nucleic acid with protein it is assumed that during the binding process a redistribution of already bound protein molecules over the nucleotide lattice occurs. The question arises whether this redistribution takes place via dissociation and re-association or whether a shuffling of bound ligands on the nucleotide lattice is also possible. Two different experiments indicate that this latter mechanism is very likely.

We first mention the results of the binding experiments between gene-5 protein and poly(dT), performed by Pörschke and Rauh [35] The association to moderate degrees of saturation of the polynucleotide has been simulated by Monte Carlo calculations. In this way rate constants were estimated for the association and dissociation reactions of isolated and single contiguously bound protein molecules, with the assumption that $\omega$ is dissociation-controlled. It has been found that the dissociation of cooperatively bound ligands proceeds with a time constant of about 120 s. Without a facilitating mechanism, the complete saturation of the poly(dT) would be rate-limited by this time constant of 120 s. However, in the mixing experiments of gene-5 protein with poly(dT) leading to complete saturation, the slowest time constant is only about 7 s. Therefore Pörschke and Rauh [35] conclude that a translocation mechanism, in which single contiguously bound protein molecules convert to isolated bound protein molecules, must facilitate the redistribution process.

Secondly, the results of Shimamoto and Utiyama [59] also indicate the involvement of a redistribution mechanism. This has been concluded as follows. At first the dissociation rate of a single contiguously bound gene-5 protein has been estimated from photo-cross linking experiments involving gene-5 protein - fd DNA complexes after addition of a large excess of short DNA fragments. Next, by multiplying this rate with the experimentally determined value for $\omega$, an estimation has been made of the rate constant for the dissociation of an isolated bound protein from the fd DNA complex (assuming that $\omega$ is dissociation-controlled). The rate constant thus obtained is 10 to 100 times larger than the measured value. In this case a fast rearrangement of isolated bound protein molecules along the nucleotide lattice to protein clusters can

account for the observation. Note that the direction of this rearrangement is opposite to, but does not exclude, the one proposed by Pörschke and Rauh (see above).

### 1.3g Protein residues involved in binding

To obtain further insight in the DNA binding properties of the gene-5 protein, attempts have been made to investigate the DNA binding domain, i.e. experiments have been performed to determine the structure of this domain and to establish which amino acid residues are involved in the protein - DNA interaction.

As has been pointed out in a preceding section, electrostatic interactions play an important role in the complex formation. Therefore we pay special attention to the basic amino acid residues the six lysyl, the four arginyl and the single histidyl residue of a gene-5 protein monomer. These, as well as the N-terminus, are candidates for binding to the negatively charged phosphate backbone The side chain pK's of free lysine, arginine and histidine are 10.5, 12.5 and 6 0 respectively. Therefore at physiological pH lysine and arginine will be positively charged and histidine will be neutral. With the help of proton NMR experiments it has been shown that the single histidyl group can only be protonated below pH 5.0, where the protein has lost its native structure [own observations]. This indicates that the histidyl residue is probably buried in the protein and will not be involved in intermolecular interactions. It also follows from these experiments that a denaturation occurs above pH 10.8. Because the apparent pK's of the arginyl and lysyl residues in the protein are larger than 10.8, a selective deprotonation, to detect a role of these residues in the complex stabilization, is not possible. However, proton NMR studies of Alma et al. have indicated that, after addition of $d(A)_8$ or $d(A)_{25-30}$, one or two lysyl residues and two or three arginyl residues are immobilized [75]. The earlier results of Coleman et al. [62], are in agreement with the observations concerning the arginyl resonances, but are conflicting in relation to the line broadening of the lysyl resonances as observed by Alma et al. [75]. However, these results [62] were obtained at a poor spectral resolution. In previous studies of Anderson et al. it has al-

ready been shown that acetylation of the ε-amino lysyl group with N-acetylimidazole prevents complex formation [61]. All the lysyl residues can become chemically modified in this way and can also become maleylated with maleic anhydride [70]. Therefore they are thought to be situated at the surface of the molecule. Using a different approach, Gray et al. have dimethylated the six lysyl amino groups, introducing in this way $^{13}C$ atoms [68]. Then a weaker binding to fd DNA is observed. From $^{13}C$ NMR spectra the conclusion was drawn that two lysyl residues are exposed to the solvent. At least two, but not more than four lysyl residues may be involved in the binding to $d(A)_7$, in agreement with Alma's results [75].

The changes observed in the 228 nm CD band of the gene-5 protein [29,33, 61,68], in the protein UV absorbance [29], in the protein fluorescence [references see above] and in the aromatic region of the proton NMR spectrum of the protein (see below) upon addition of viral DNA, all indicate a stacking interaction between some of the protein tyrosyl residues and the nucleotide bases. Alma et al. have located the resonances of the five tyrosyl and three phenylalanyl residues in the aromatic region of the proton NMR spectrum [76]. Upfield proton NMR shifts [60,62,67,76-79] and one dimensional nuclear Overhauser effects [69,76] indicate that two tyrosyl residues and one phenylalanyl residue are probably involved in the binding. In experiments of Anderson et al. three out of the five resonances in the $^{19}F$ NMR spectrum of m-fluorotyrosyl labeled gene-5 protein are found at a position which is expected for free m-fluorotyrosyl, whereas the two other resonances are shifted downfield. This indicates that three tyrosyl residues are situated at the protein surface and two are buried in the protein inner structure [61]. From peptide mapping it has been concluded that the three tyrosyl residues at the surface are situated at the positions 26, 41 and 56 in the gene-5 protein primary structure [61]. Coleman et al. have shown that upon addition of tetranucleotides or octanucleotides the two downfield $^{19}F$ resonances do not undergo a detectable shift, while at least two of the tyrosyl residues at the surface show a small upfield shift [62,77]. Furthermore, the three tyrosyl residues at the surface can become acetylated with N-acetylimidazole which prevents complex formation, or can become nitrated by tetranitromethane thereby strongly weakening the binding affinity [61]. On the other hand the protein is protected against nitration after binding to fd DNA. Evidence for

the existence of three surface tyrosyl residues has also been obtained from photo-CIDNP experiments [80,81]. In complexes with the tetranucleotide d(pCGCG) the hydroxyl group of these tyrosines is no longer accessible to the excited flavine dye, due to shielding and (or) hydrogen bonding [81], and their photo-CIDNP signal disappears. All the experimental results described above, indicate a stacking interaction between the nucleic acid bases and two tyrosyl residues and one phenylalanyl residue.

By irradiation of the in vitro complex [82] or the in vivo complex [64] the protein can become covalently linked to fd DNA. This cross-linking to DNA via the sulfhydryl group of cysteine provides evidence for the positioning of this residue in the protein - DNA interaction domain. The cysteine in the free protein can be alkylated [64,82] and mercurated [61], but does not react with the bulky Ellman's reagent [61]. This supports the idea that although the cysteine is situated at the protein outer surface it is not completely exposed to the solvent. Consistent with these experiments are the results from X-ray analysis of gene-5 protein crystals. From these the conclusion was drawn that the cysteine residue is buried in the hydrofobic core beneath the central three-stranded β-sheet, and is centered in the proposed DNA binding cleft [41].

Finally, it is interesting to ask to what extent the gene-5 protein structure may change upon binding. Several features, characteristic for a folded protein structure observed in the proton NMR spectrum of the protein are also present in proton NMR spectra of gene-5 protein - nucleic acid complexes [37]. This holds for the resonance position of the lowfield shifted α protons and for the overlap of methyl resonances at about 0.5 ppm. The small, but important changes observed in the protein NMR spectrum upon complexation are assigned to residues which are situated in or near the DNA binding domain. Therefore we think that, except for structural changes in this DNA binding domain, only minor conformational changes take place and that the overall secondary structure of the protein is conserved upon binding.

## 1.4 OUTLINE OF THIS THESIS

In the preceding paragraphs a review of the literature concerning the interaction of gene-5 protein with nucleic acids was given, which can serve as an introduction of the investigations presented in this thesis. First, a brief outline of the studies described in het following chapters will be given.

It is not known so far whether any nucleotide or sugar specificity plays a role in the binding of the gene-5 protein to nucleic acids. Once this problem has been resolved it may become possible to formulate general rules concerning the specificity of the binding of single stranded DNA binding proteins after comparison of the properties of several of these proteins For example, it is necessary to know the binding affinity of the gene-5 protein to poly(dA) and poly(dU) to give an answer to the problem whether the different binding affinities of the gene-5 protein and the T4 encoded gene-32 protein to A and U stretches form an explanation for the results obtained from the experiments of Fulford and Model [83], from which it can be concluded that the proteins cannot replace each other in their translational repressor functions, see also section 1.2. Furthermore, it is still not clear whether there exists a difference in the salt dependency of the binding of gene-5 protein to different nucleic acids. After determination of the binding affinities of the gene-5 protein to various polynucleotides possible implications for an understanding of its role in vivo can be discussed. The subjects mentioned above are treated in Chapter 3.

On comparison of the binding of gene-5 protein to polynucleotides and oligonucleotides the question was put forward whether the difference in binding stoichiometry found for these two types of nucleic acid would indeed arise from the difference in chain length or would be a result of the different experimental techniques used and therefore of a disparity in protein concentration [84]. In other words, a protein concentration dependency of the binding properties was suggested. The answer will be given in Chapter 4.

The properties of the gene-5 protein - oligonucleotide binding mode, like binding cooperativity, binding strength and salt dependency of the binding are studied in the Chapters 4 and 6. Also a comparison of these properties to

those characteristic for the binding of gene-5 protein to polynucleotides is made. The various results discussed in these chapters are consistent with the already mentioned low value of the cooperativity parameter for the binding of gene-5 protein to oligonucleotides. The difference in binding to oligo- and polynucleotides is also reflected in a different mechanism for the dissociation reaction of the complexes (confer Chapters 5 and 6). The dual binding mode of the gene-5 protein - polynucleotide complex at low salt concentration is one of the items of Chapter 6 as well. Here also experimental evidence is presented for the assumption that in the oligonucleotide binding mode the two dimer halves of the gene-5 protein are bound to the same strand.

Finally, the NMR experiments in Chapter 7 are carried out to detect a possible change in the DNA conformation upon complex formation.

Before we turn to the experimental results (Chapters 3 and following Chapters) some underlying theoretical considerations will be presented in Chapter 2.

# CHAPTER 2

# THEORETICAL ASPECTS

## 2.1 BINDING THEORIES

Several theories are in current use to describe the binding of a ligand to a one-dimensional lattice [71,85-89]. Most of these theories were developed for a special application and their suitability depends on the length of the lattice (polymer or oligomer), on the number of binding modes (a distinct binding mode is characterized by a definite structure of the ligand-lattice complex), on the number of ligand binding contacts (single or multiple contact binding) and on the condition of sequence specific or non-specific ligand binding. In the next sections we will only describe theories which deal with sequence non-specific binding and with multiple contact binding, i.e. one ligand may cover a site consisting of more than one single lattice residue and thereby exclude a number of potential binding sites from binding. To describe the nucleic acid binding properties of the gene-5 protein it is sufficient to treat three theories: single mode binding to polymers (section 2.1a), single mode binding to oligomers (section 2.1b) and multiple mode binding to polymers (section 2.1c). The following notations and assumptions will be used:

- The lattice consists of a linear strand of M residues.
- $[P_f]$ is the concentration of ligand free in solution.
- The ligand can bind in m different binding modes, each labeled with an index i.
- A bound ligand covers exactly $n_i$ consecutive lattice residues; the parameter n is called the binding stoichiometry.
- Binding of a ligand to an isolated site takes place with an association constant $K_{int,i}$, called the intrinsic binding constant of mode i.
- Interactions between bound ligands only occurs between nearest neigbours. Binding of a ligand in mode i next to an already bound ligand in mode i' (i' may be i) is a factor, $\omega_{i,i'}$ (the cooperativity factor), stronger than binding

to an isolated site. For non-interacting ligands $\omega_{1,1'} = 1$. For positive cooperativity $\omega_{1,1'}$ is greater than 1 and for negative cooperativity $\omega_{1,1'}$ lies between 0 and 1.
- The binding density, $\nu_1$, is the ratio of the number of ligands bound in mode 1 to the total number of lattice residues. The degree of binding, or degree of saturation, $\alpha_1$, equals $n_1\nu_1$ and lies between 0 and 1

In each of the theories described below mathematical expressions are presented, which make it possible to calculate the binding density as a function of $[P_f]$. Via simulation procedures of the experimental results with these calculations values for the binding constants and stoichiometries can be obtained.

For a detailed description of the various binding theories we refer to the original papers, mentioned in the next sections In this chapter only abstracts will be given, sufficient for the use and understanding of the final equations.

## 2.1a Single mode binding to polymers

The theory of multiple contact binding of a ligand, which binds in only one mode to a linear polymer, has been developed by McGhee and Von Hippel [71] and by Schwarz [86]. We will follow the description of the first authors, which leads to closed analytical expressions for the binding curves. In the description of the binding of ligands covering more than one lattice residue it must be reckoned that the number of binding sites is not only dependent on the degree of saturation, but also on the distribution of the complexed ligands. On a partial saturated lattice the number of still available binding sites is determined by the number of gaps and their size. A gap of g lattice residues contains g-n+1 binding sites as long as $g \geq n$; for $g < n$ the number of binding sites is zero. To be able to count the number of available binding sites, the probability $P_g$ is introduced, which is the probability of finding a gap of exactly g free lattice residues in a row. A convenient approach to the determination of $P_g$ makes use of the so-called conditional probabilities.

The conditional probability $(r_1r_2)$ is the probability of finding a (polymer) residue in state $r_2$ while the residue on the left is in state $r_1$. The following expressions for the conditional probabilities can be derived:

$(b_nb_1)$ = the conditional probability that the right end of a bound ligand is adjacent to the left end of another bound ligand

$$= (1 - (n - 2\omega + 1)\nu - R) / (2\nu(\omega - 1)) \qquad (2.1)$$

and

$(ff)$ = the conditional probability that a free residue will be at the left side adjacent of a free residue

$$= ((2\omega - 1)(1 - \alpha) + \nu - R) / (2(\omega - 1)(1 - \alpha)) \qquad (2.2)$$

and analogously $(b_nf) = 1 - (b_nb_1)$, $(fb_1) = 1 - (ff)$, in which

$$R = \{(1 - \nu(n + 1))^2 + 4\omega\nu(1 - \alpha)\}^{1/2} \qquad (2.3)$$

The average number of ligand binding sites per gap equals

$$s = \Sigma_g\ (g-n+1)\ P_g \qquad (g = n,\ n+1,\ ..\ M) \qquad (2.4)$$

in which

$$P_g = (b_nf)(ff)^{g-1}(fb_1) \qquad (2.5)$$

There are three types of binding sites: isolated, single contiguous and double contiguous, which means that upon binding of a ligand to these sites 0, 1, respectively 2 ligand-ligand contacts will be formed. To derive an over-all binding equation we have to consider three binding equilibria, i.e. equilibria for binding of a ligand to each of the three different binding sites. The equations of the binding constants are:

$$K_{int} = \text{[average number of isolated bound ligand per lattice]} / [P_f]\text{[average number of free isolated sites per lattice]} \quad (2.6)$$

$$K_{int}{}^{\omega} = \text{[average number of single contiguous bound ligand per lattice]} / [P_f]\text{[average number of free single contiguous sites per lattice]} \quad (2.7)$$

and

$$K_{int}{}^{\omega^2} = \text{[average number of double contiguous bound ligand per lattice]} / [P_f]\text{[average number of free double contiguous sites per lattice]} \quad (2.8)$$

If B is the average number of ligands bound to a lattice, then the number of gaps ($\geq 0$) equals B+1 and the average number of each of the three types of sites per lattice (as appears in the denominators of Eqs. 2.6-2.8) is given by the product of (B+1) and the average number of ligand binding sites of the given type, which can be derived by using the probability $P_g$, given in Eq. 2.5. Rearranging the Eqs. 2.6-2.8 and summing for the number of bound ligand per lattice (which equals B) one gets the over-all binding equation for infinitely long lattices [71]:

$$\nu/[P_f] = K_{int}(1 - \alpha) \cdot (((2\omega - 1)(1 - \alpha) + \nu - R) / (2(\omega - 1)(1 - \alpha)))^{n-1}$$
$$\cdot((1 - (n+1)\nu + R) / (2(1 - \alpha)))^2 \qquad (2.9)$$

which equals the final equation given by McGhee and Von Hippel, apart from a sign error. The degree of saturation $\alpha$ (which equals $n\nu$) of a polymer is only a function of $\omega$, n and the product $K_{int}[P_f]$, as can be concluded from an inspection of Eq. 2.9. Note that it is independent of the total polymer or total ligand concentration. It is illustrative to consider the influence of the magnitude of the cooperativity factor on the shape of binding isotherms. In Fig. 2.1 binding isotherms are shown for a ligand binding with a stoichiometry of n = 4 to a polymer of infinite length as a function of the cooperativity factor. First we consider the curves at low degrees of saturation. At a particular value of $[P_f]$ and of $K_{int}\omega[P_f]$ stronger binding is observed for lower values of $\omega$, since plotting of the data in this manner means that for larger values of $\omega$, $K_{int}$ becomes smaller. At low degrees of saturation ligands will only be bound isolated (with a binding constant per bound ligand of $K_{int}$) and therefore stronger binding will be observed. Next we consider the point at half saturation (see Fig. 2.1). It follows from Eq. 2.9 that at this point for high values of $\omega$ relative to n, the reciprocal value of the free ligand concentration equals the product $K_{int}\omega$:

$$[P_f]^{-1} = K_{int}\omega \qquad (\text{if } \omega >> n) \qquad (2.10)$$

One can use this relation to estimate the value of the effective binding constant ($K_{int}\omega$) from binding isotherms. Finally steep curves, resembling all or none binding, are obtained at very high values of the cooperativity parameter (see Fig. 2.1). The conversion from free to totally saturated polymer is obtained for only a small increase in the total (and therefore in the free) ligand concentration above a minimum threshold value, which depends on the magnitude of $K_{int}$.

It can be shown that the effect of higher values of the binding stoichiometry on the degree of saturation is similar to the effect of lowering the cooperativ-

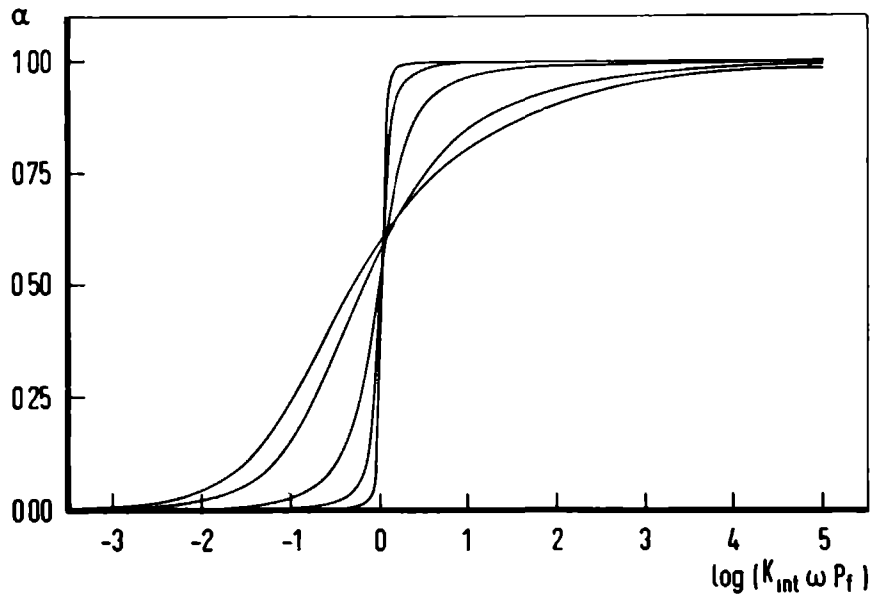



Fig 2 1 Influence of the cooperativity parameter on ligand - polymer binding The ligand covers four lattice residues (n = 4) and the polymer has an infinite length The degree of saturation of the polymer, α, is given as a function of the product of $K_{int}\omega[P_f]$ for different values of the cooperativity constant and is obtained by using Eq 2 9 ω is taken 1, 2, 20, 200 and 2000 for the curves in the order of increasing steepness

ity factor [90]. This can be understood as follows: many lattice residues are excluded from binding due to binding site overlap if n is large

For finding the concentrations at equilibrium of each of the three types of bound ligand, i.e ligand bound isolated, single contiguously or double contiguously the next procedure can be followed. At first a relation for the concentration at equilibrium of clusters of q consecutive bound ligands, $[CL_q]$,

is given. $[CL_q]$ equals the total concentration of lattice residues, $[N_{tot}]$, times the joint probability for a stretch of q consecutive bound ligands:

$$[CL_q] = [N_{tot}] \{f\} (fb_1)(b_n b_1)^{q-1} (b_n f) = [P_b] (1 - p)^2 p^{q-1} \qquad (2.11)$$
$$p = (b_n b_1)$$

where $\{f\} = 1 - \alpha$ is the probability for a free residue while the conditional probabilities in Eq 2.11 are given in Eqs. 2.1-2.3. $[P_b]$ is the concentration of bound ligands. The concentration at equilibrium of each of the three types of bound ligand can now be given, using Eq. 2.11. For the concentration of isolated bound ligands it follows that

$$[P_i] = [CL_1] = [P_b] (1 - p)^2 \qquad (2.12)$$

Every cluster of two or more ligands possesses a single contiguously bound ligand at both ends. Therefore at equilibrium the concentration of single contiguously bound ligands is

$$[P_s] = 2 \sum_{q=2}^{\infty} [CL_q] = 2[P_b] (1 - p)^2 \sum_{q=2}^{\infty} p^{q-1} = 2[P_b] (p - p^2) \qquad (2.13)$$
$$(0 \leq p \leq 1)$$

Finally, the concentration of double contiguously bound ligands at equilibrium amounts

$$[P_d] = [P_b] - [P_i] - [P_s] = [P_b]p^2 \qquad (2.14)$$

Eqs. 2.11-2.14 have been used to construct Fig. 2.2, in which the probability

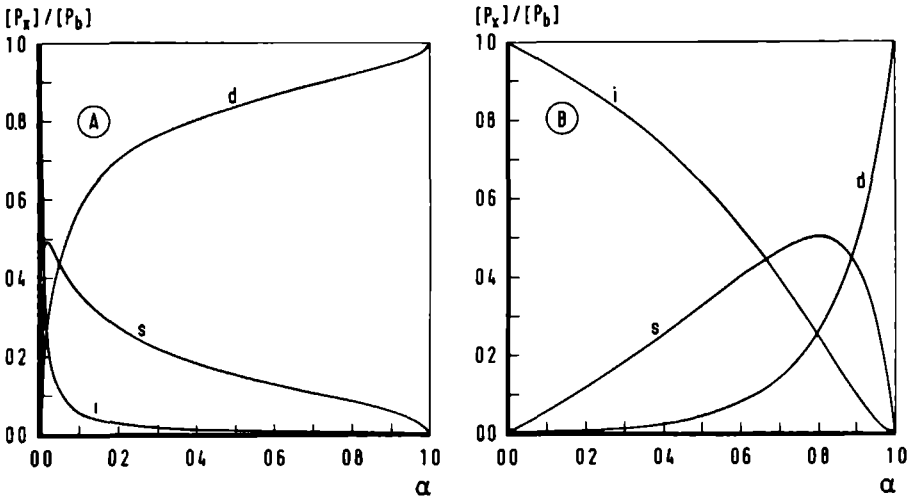



Fig. 2.2. Distribution of ligands bound to an infinitely long polymer in the isolated (x = i), single contiguous (x = s) or double contiguous (x = d) fashion, as a function of the degree of saturation, $\alpha$, of the polymer. $[P_x]$ is the concentration of ligand bound in each of the different modes and $[P_b]$ is the total concentration of bound ligand. Curves are given for binding of a ligand with n = 4 and with $\omega$ = 500 (A) or $\omega$ = 1 (B).

that a bound ligand is bound isolated, single contiguously or double contiguously is shown for two cases. First (Fig. 2.2A) binding parameters are used which are thought to be representative for the gene-5 protein polynucleotide binding: n = 4 and $\omega$ = 500, in which case the expression for $p = (b_n b_1)$ is given in Eq. 2.1. Secondly, the distribution is calculated for non-cooperative binding, in which $p = \nu/(1 - (n-1)\nu)$ [71], $\omega$ = 1 and n is taken 4 (Fig. 2.2B). For non-interacting ligands the probability $[P_i] / [P_b]$, i.e. the probability that a bound ligand is bound isolated (note that this is not equal to the probability that a ligand is bound isolated) roughly equals $(1 - \alpha)$, the

fraction of the lattice residues which are not covered with ligand. There is a maximum at 80 % saturation for the probability that a bound ligand is bound single contiguously. The number of double contiguously bound ligands increases at higher degrees of binding. In the cooperative case (Fig. 2.2A) the distribution of bound ligands in the three different forms is not only ruled by entropy but also by the counteracting ligand-ligand cooperativity. At saturation of the lattice of 5 % or more the percentage of bound ligands which are bound isolated is always lower than 10 %. Because of the cooperativity binding is favoured for ligands at sites contiguously to ligands already bound relative to binding to isolated sites. The probability that a bound ligand is bound single contiguously again shows a maximum (Fig 2.2A), but at much lower degrees of binding than for the non-cooperative binding (Fig. 2.2B).

Fig. 2.3 shows the distribution of protein clusters of different length for two examples of cooperative binding. The curves indicate the percentage of bound ligand at equilibrium in clusters up to the length N (for which we take as examples N = 1, 2, 5, 10, 15, 20, 30, 40, 100, 200 or infinite) as a function of the degree of saturation of the polymer. Curves are given for n = 4 and $\omega$ = 500 (left) and n = 7 and $\omega$ = 5000 (right), which values are representative for the binding to polynucleotides of respectively gene-5 protein [this thesis] and gene-32 protein [91]. The indicated percentage can be derived from Eq. 2.11 and equals

$$\text{percentage of bound ligand in clusters of N ligands or in smaller clusters} = (\Sigma_{q=1..N}\, q.CL_q \,/\, \Sigma_{q=1..\infty}\, q.CL_q)\ .\ 100\ \%$$

$$= (1 - p)^2\ \Sigma_{q=1..N}\, q.p^{q-1}.100\ \% \qquad (2.15)$$

The average cluster length increases for higher degrees of saturation. From Fig. 2.3 (left) it can be estimated for example that at half saturation of a polynucleotide with gene-5 protein one half of the clusters would only contain less than 17 ligands, or 90 % of the clusters is smaller than about 40 ligands at this degree of saturation. For the stronger cooperative complexation of

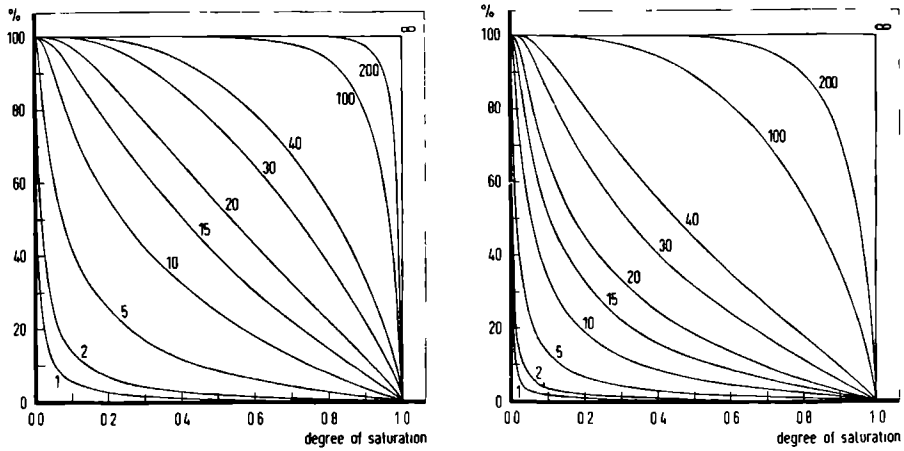



Fig 2 3 Distribution of ligand clusters bound to an infinitely long polymer The percentage of bound ligand bound in clusters of N ligands or in smaller clusters as a function of the degree of saturation (see Eq 2 15) is given The numbers N are indicated along the various curves (1, 2, 5, etcetera) Curves are given for n = 4 and ω = 500 (left) and n = 7 and ω = 5000 (right)

gene-32 protein these values are more than two times larger (Fig. 2.3 right). We remark that electron microscopic determination of the cluster length distribution, as given in Fig. 2.3, can provide an estimation of the cooperativity factor.

### 2.1b Single mode binding to oligomers

When a lattice can no longer be considered infinitely long, the theory of McGhee and Von Hippel (see above) cannot be applied. For this case one can have recourse to a theory developed by Epstein for the binding of ligands to a finite one-dimensional lattice [87]. This theory leads to the following relation between the free protein concentration, $[P_f]$, and the binding density $\nu$:

$$\nu = (1/M)\ \Sigma_q\ \Sigma_j\ q\ \sigma(M,n,q,j)\ (K_{int}[P_f])^q\omega^j\ /\ \Sigma_q\ \Sigma_j\ \sigma(M,n,q,j)(K_{int}[P_f])^q\omega^j$$

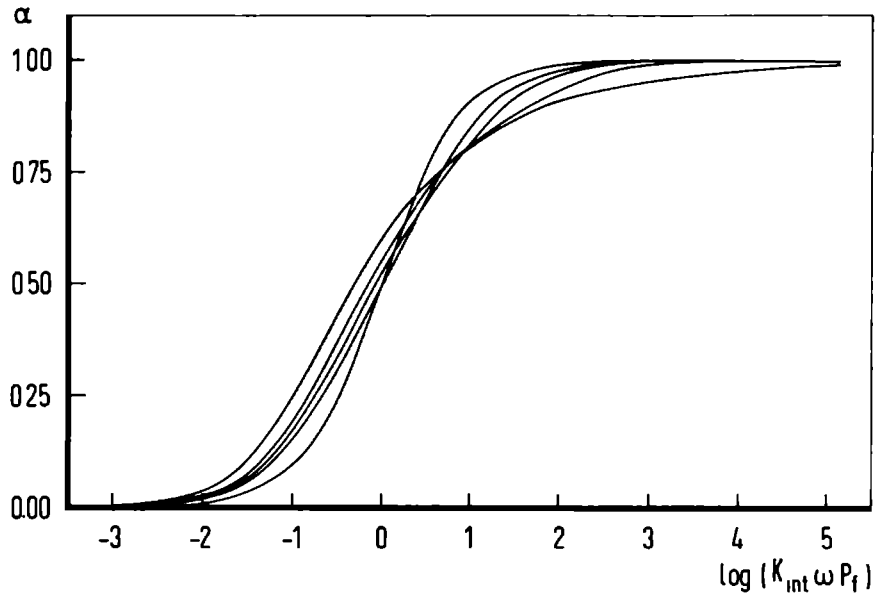
$$(q = 0,1,\ ..\ g,\ j = 0,1,\ ..\ (q-1)) \qquad (2.16)$$

in which the maximum number of ligands which may bind to the lattice is denoted by g, the highest integer less than or equal to the ratio M/n. The configuration entropy factor

$$\sigma(M,n,q,j) = \binom{M-nq+1}{q-j}\ \binom{q-1}{j} \qquad (2.17)$$

gives the number of ways q ligands can bind to a lattice of M residues, each occupying n consecutive lattice residues and forming j ligand-ligand contacts. It follows from Eq. 2.17 that the binding density is only a function of the lattice length M, the stoichiometry n, the cooperativity factor $\omega$ and the product $K_{int}[P_f]$. Note that it is independent of the total oligomer concentration or total ligand concentration.

In Fig. 2.4 the influence of binding cooperativity and lattice length on the degree of saturation $\alpha$ are shown, calculated using Eq. 2.17. In the lower part of Fig. 2.4 binding isotherms are given for the binding of a ligand to an octadecamer with a stoichiometry of n = 3. As for binding isotherms for the binding to polymers (see Fig. 2.1) steeper curves are obtained at higher values of $\omega$. However the absence of any common intersection point for the various curves precludes the derivation of a binding constant from the free ligand concentration as could be done for binding to polymers at half saturation (see Eq. 2.9). In the upper part of Fig. 2.4 binding isotherms are given for the non-cooperative binding of a ligand with a stoichiometry of n = 3, to lattices of increasing length. At low degrees of binding stronger binding is observed for longer lattice lengths, because of the increased number of potential binding sites. At high degrees of binding however longer lattices will be less saturated because there is an increase in the number of residues which are excluded from binding due to the existence of small


α
100
075
050
025
0.00
-3
-2
-1
0
1
2
3
4
5
log (K int ω Pf )

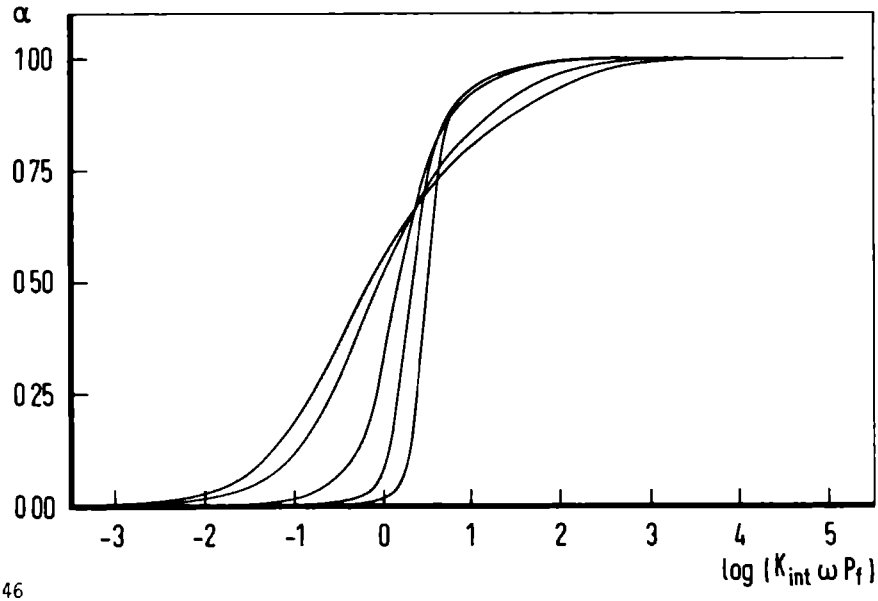

α
100
075
050
025
0 00
-3
-2
-1
0
1
2
3
4
5
log (K int ω Pf )

Fig 2 4 Binding isotherms for the binding of a ligand to an oligomer with n = 3 The curves are obtained using Eq 2 16 The degree of saturation is given as a function of the product $K_{int}\omega[P_f]$ In the upper part of the figure the influence of the lattice length on the degree of saturation for non-cooperative binding is indicated For $\log(K_{int}\omega[P_f]) < 0$ steeper curves are obtained in the order M = ∞ (polymer), 18, 12, 6 and 3 In the lower part of the figure the influence of the cooperativity parameter is shown for binding to an octadecamer (M = 18) Steeper curves are obtained for ω increasing in the range 1, 2, 20, 200 respectively 2000

stretches of empty lattice residues, which are smaller than n. The existence of these empty lattice residues is favoured by the increase in entropy. The strongest binding will now be observed to a lattice for which these stretches cannot exist, i.e. for binding to a lattice for which M equals n. This is demonstrated in Fig. 2.4.

2.1c Multiple mode binding to polymers

The multiple mode binding of ligands to polymers is described by Schwartz and Stankowski [89]. Distinct binding modes are characterized by an inequivalence of one or more of the following properties: the number of lattice residues covered per ligand, the intrinsic binding constant and the cooperativity of the binding process. As in the derivation of McGhee and Von Hippel (see before), binding equations are expressed in terms of the equilibrium constants, the free ligand concentration and conditional probabilities. The conditional probability of finding a bound ligand in mode i next to the right of a free residue is denoted by $(fb_{1,i})$ At m binding modes a set of m linear equations for the parameters $(fb_{1,i})$ is obtained:

$$\omega_{i,i}\,\{(1/(ff))^{n_i} / K_{int,i}[P_{f,i}]\}\,(fb_{1,i}) = 1 + \Sigma_{i'}\,(\omega_{i,i'} - 1)\,(fb_{1,i'})$$

$$(i,i' = 1,2, \ldots m) \qquad (2.18)$$

By means of Cramer's rule it follows that

$$(fb_{1,i}) = Det_i[\ (ff)\ ]\ /\ Det[\ (ff)\ ] \qquad (2\ 19)$$

where the Det-terms stand for the appropriate determinants. Because (ff) equals $1 - \Sigma_i\ (fb_{1,i})$ for $i = 1$ to m, it follows that

$$(1 - [\ 1\ /\ (ff)\ ]\ Det\ [\ (ff)\ ] + [\ 1/(ff)\ ] \sum_{i=1}^{m} Det_i\ [\ (ff)\ ] = 0 \qquad (2.20)$$

Only the smallest root for (ff) satisfies the condition $0 \leq [(ff), (fb_i)] \leq 1$ Once this root has been found, the parameters $(fb_i)$ can be solved from Eq 2.19. Consider the cooperativity parameter to be independent of the relative sequence of the two cooperatively bound ligands: $\omega_{i,i'} = \omega_{i',i}$. Finally the expression for the binding density in mode i, $\nu_i$, is obtained:

$$\nu_i = (fb_{1,i}) \sum_{i=1}^{m} [1 + (\omega_{i,i} - 1)\ (fb_{1,i})] \qquad (2.21)$$

The calculation of the binding density as a function of the binding stoichiometries, the binding constants, the cooperativity parameters and the free ligand concentration(s) is somewhat more complicated as in the case of single mode binding (Eq. 2.9). Now each time a higher order equation has to be solved to find the right root for (ff).

There exists a correspondence between the effects of a higher positive cooperativity, which causes an increase in binding strength, and a smaller binding stoichiometry, which enhances the amount of bound protein because of a decrease in the number of unavailable overlapping binding sites This has been demonstrated for single mode binding [90]. Furthermore, simulations of binding curves for mutually non-cooperative binding modes show that at suffi-

cient high products of the free ligand concentration and the intrinsic binding constant, binding always takes place in the mode with the smallest number of contacts per ligand, irrespective of the value of the cooperativity parameters. An example of this phenomenon forms the dual mode binding of gene-5 protein to polynucleotides at low ionic strength in Chapter 6, which is simulated by using Eqs. 2.18-2.21.

## 2.2 KINETIC THEORIES

The gene-5 protein plays its role at continuously changing conditions in the cell. For a proper understanding of its function also knowledge of the kinetic properties is needed. The various rate constants for binding of a ligand that binds cooperatively to a linear lattice are shown in Fig 2.5. This figure represents the binding of a ligand P to a lattice S, which results in an isolated (i), single contiguously (s) or double contiguously (d) bound ligand. A rearrangement from an isolated bound ligand to a single contiguously bound ligand and vice versa may take place (not indicated in Fig. 2.5). The cooperativity parameter $\omega$ is assumed to consist of an association (a) and a dissociation (d) part:

$$\omega = \omega_a \cdot \omega_d \tag{2.22}$$

From the definitions stated in Fig. 2.5 and Eqs. 2.6 - 2.8 it follows that

$$k_d = k_s/\omega_d = k_i/\omega_d^2; \quad k_{d'} = k_{s'} \cdot \omega_a = k_{i'} \cdot \omega_a^2 \tag{2.23}$$

If $\omega_d$ equals unity, the cooperativity of ligand binding originates from the association process and the recognition of newly bound ligands must play an important role. However, if $\omega_a$ equals unity, the cooperativity is associated with the ligand dissociation and a conformational change after initial binding, increasing the ligand-ligand interaction, might have taken place. Experimental

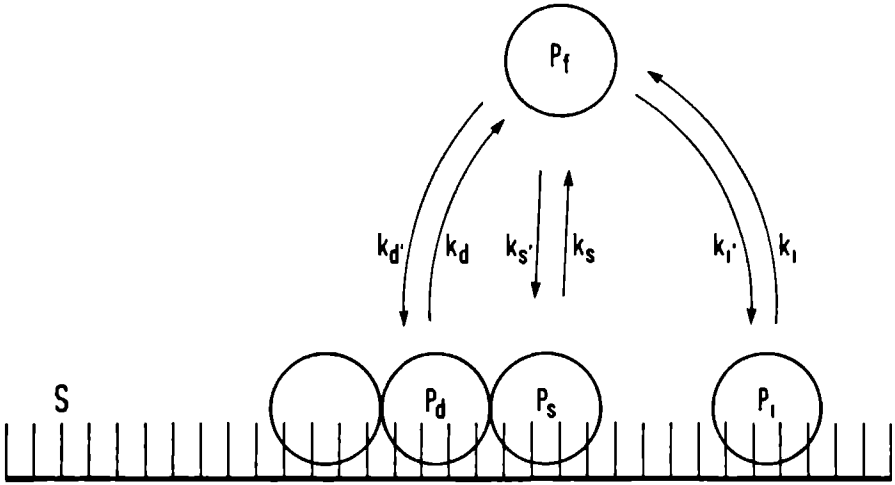



Fig 2 5 Reaction scheme in which the association and dissociation of a ligand P is given, which can bind in three different ways to the polymer S, i e to an isolated (i), a single contiguous (s) or a double contiguous (d) site $P_f$ denotes the free protein Association reaction constants are labeled with an accent

results of the gene-5 protein and the gene-32 protein (encoded by bacteriophage T4) kinetics [35,59,92] show that binding of ligands to an isolated site takes place at an association rate which is expected for a diffusion-controlled reaction. Cooperative binding of ligands can never exceed this upper limit. For this reason cooperativity must be linked to the dissociation in these cases and from now on we take $\omega_a = 1$ and $\omega = \omega_d$

Furthermore, any rearrangement like translocation without dissociation (sliding), translocation with dissociation (hopping), direct transfer [92], etcetera, can facilitate an efficient and fast distribution of the ligand.

A study of the association kinetics has the disadvantage of the great complexity of both the association, dissociation and redistributing processes. However, valuable experimental and theoretical results have been obtained [35,59,92,93]. We will limit ourselves to a description of the irreversible dissociation kinetics for which sound theoretical models have been worked out. In the case of protein nucleic acid reactions the validity of these models can be checked by instantly increasing the ionic strength of a protein nucleic acid sample at equilibrium to such an extent that association is strongly weakened and can be neglected.

## 2.2a The non-redistributing model

In the non-redistributing model irreversible dissociation of ligands takes place from an infinite lattice. During the dissociation process no rearrangement of the bound ligands is allowed. This model was first introduced by Lohman [94]. It is assumed that for strong cooperative binding the dissociation of double contiguously bound ligands can be neglected in comparison to the dissociation of single contiguously and isolated bound ligands and even that dissociation only takes place from the ends of a cluster ('peeling'). For lower values of the binding cooperativity, dissociation of proteins bound in between two other proteins becomes important as well and this situation has been treated by Balazs and Epstein [95]. Hence we consider the Lohman model because high cooperativity is often observed in protein - polynucleotide reactions. This makes the solution of the rate equations much simpler. We seek for a general equation to which experimental data can be fitted. First we denote the concentration of clusters consisting of q consecutively bound ligands by $[CL_q]$. The time dependent change of the concentration of bound ligands, $[P_b]$, is given by

$$(d/dt)\,[P_b] = -\,k_i[CL_1] - 4\,k_s\,[CL_2] - 2k_s \sum_{q=3}^{\infty} [CL_q] \qquad (2.24)$$

A cluster of more than one ligand has two ends from which dissociation can take place, which gives rise to the factor of two in the third right hand term of Eq. 2.24. A cluster of two ligands is supposed to dissociate as a whole, thereby extra doubling the dissociation rate (second right hand term). By combination of Eq. 2.24 and the Eqs. 2.12 - 2.14 it can be shown that at an initial incomplete saturation integration leads to a sum of two exponentials for the time dependence of the concentration of bound ligand:

$$[P_b(t)] \,/\, [P_b(0)] = (1 - p(0))^2\, e^{-k_i t} + p(0)(2 - p(0))\, e^{-2(1 - p(0))k_s t} \qquad (2.25)$$

The first term originates from the dissociation of isolated bound ligands, which is fast compared to the dissociation from cluster-ends in the second term. However, at moderate degrees of saturation and high values of $\omega$ the influence of the dissociation of isolated bound ligand can be neglected because the fraction of ligands bound at isolated sites is very low. The apparent rate constant equals $2k_s(1 - p(0))$ in this case. This situation has also been treated by Schwartz and Watanabe [90], whose rate equation (44a) equals the second right hand term of Eq. 2.25 presented in a somewhat different form.

When at $t = 0$ complete saturation ($p(0) = 1$) is obtained, the solution of the rate equations leads to a zero order kinetics [94]:

$$[P_b(t)] = [P_b(0)] - 2k_s[S(0)]t \qquad (2.26)$$

where $[S(0)]$ denotes the total concentration of nucleotide strands at $t = 0$. The number of cluster-ends stays constant as long as the clusters are large. However, towards the end of the dissociation the concentration of ends will begin to decrease and a constant rate of dissociation is no longer observed.

For the situation in which $\omega_a = 1$ and $\omega = \omega_d$ (see Eq. 2.22) the dissociation constants $k_i$, $k_s$ and $k_d$ are simply related (Eq. 2.23) and the rate process,

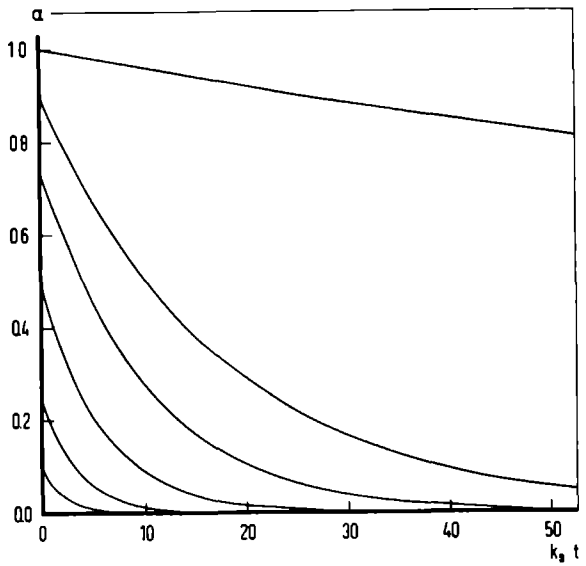



Fig 2 6 Dissociation curves for the non-redistribution dissociation model, given by Lohman [94], see section 2 2a The curves have been calculated using Eqs 2 25 and 2 26 and using n = 4 and ω = 500 Curves are given for various initial degrees of saturation (α(t=0)) and they depend on the product of the reaction rate for dissociation of a protein from a cluster end ($k_s$) and time (t)

as determined by the decrease of the ratio $[P_b(t)] / [P_b(0)]$ with $k_s t$, is only dependent on the parameters ν(0) (initial binding ratio), n and ω. In Fig. 2.6 an example of such a dissociation process is given for n = 4 and ω = 500, calculated with Eqs. 2.25, 2.26 and the relation

$$\alpha(t) = \alpha(0) \; . \; [P_b(t)] / [P_b(0)] \qquad (2.27)$$

Because the distribution of bound ligands in the isolated, the single contiguous and the double contiguous way is strongly dependent on the degree of saturation of the polymer (see section 2.1a and Figs. 2.2A and 2.3) very different dissociation time curves are obtained at different initial degrees of saturation, see Fig. 2.6. Dissociation is completed for lower initial degrees of saturation for two reasons. First, the fraction of the fastest dissociating ligands, i.e. ligands bound isolated, is larger at lower degrees of binding.

Furthermore, after some dissociation the 'pool' of these isolated bound ligands is empty and the effective first order dissociation rate constant decreases from $k_l$ to $2(1 - p(0))k_s$ (see Eq. 2.25), which has a greater value at lower initial degrees of binding. On the other hand, at initial complete saturation zero order binding is observed (see Eq. 2.26). Since in this case no isolated bound ligands are present at t = 0, the initial dissociation rate is a factor $\omega$ slower than for the other curves in Fig. 2.6.

Experimental results for the gene-32 protein - nucleic acid dissociation can be quantitatively described by Eqs. 2.25 and 2.26 [84,94,96]. Especially, the observation of zero-order dissociation kinetics (when p(0) = 1) was considered [94] to provide qualitative evidence for a mechanism in which the major mode of dissociation is via singly contiguously bound ligands while no rearrangement of clusters occurs during the dissociation. However, this will be commented in Chapter 5.

### 2.2b The immediately-redistributing model

In sharp contrast to the previous model in this model it is assumed that during the dissociation the ligand clusters on the polymer lattice immediately redistribute themselves to form a new equilibrium according to the Eqs. given by McGhee and Von Hippel (see section 2.1). The redistribution rate constants are large compared to the other rate constants. This means that the overall dissociation rate is independent of the redistribution rate. The time dependency of the concentration of bound ligands equals (see Fig. 2.5 for definitions)

$$(d/dt)\ [P_b] = -k_l[P_l(t)] - k_s[P_s(t)] - k_d[P_d(t)] \qquad (2.28)$$

The concentrations of the three types of bound ligands are given in Eqs. 2.12-2.14, while all terms are time-dependent now. Eq. 2.1 contains the expression for the parameter $p(t) = (b_n b_1)(t)$ in Eqs. 2.12-2.14, as a function of n, $\omega$ and $\nu$. The combination of Eqs. 2.12-2.14 and Eq. 2.28 leads to the expression for the time-dependency of the concentration of bound ligands

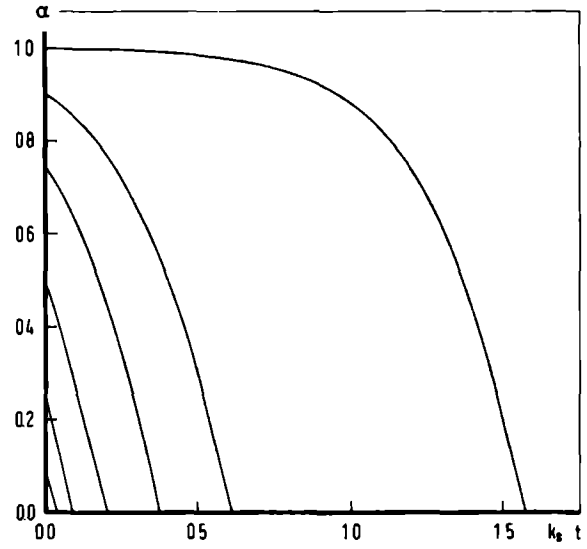



Fig 2 7 Dissociation curves for the Instantaneous Establishment of Lattice Equilibrium dissociation model, given by Epstein, see section 2 2b The curves have been calculated using Eqs 2 1 and 2 29 in which n = 4 and ω = 500 Curves are given for various initial degrees of saturation (α(t=0)) and they depend on the product of the reaction rate for dissociation of a protein from a cluster end ($k_s$) and time (t)

$$(d/dt)\ [P_b] = -\ [P_b(t)]\ \{(1-p(t))^2 k_i + 2(p(t) - p(t)^2)k_s + p(t)^2 k_d\} \qquad (2.29)$$

At known values of n, ω and ν(t = 0) the time dependence of α (see Eq. 2.26) can now be obtained by using Eq. 2.1 for the time-dependency of p(t) followed by numerically integration of Eq. 2.29. In Fig 2.7 curves are shown for various initial degrees of binding, n = 4 and ω = 500. Dissociation proceeds roughly one or two orders of magnitude faster than in the non-redistributing case (compare to Fig. 2.6). Again, the lower the initial degree of binding, the higher the initial fraction of isolated bound ligands and the faster the fraction of bound ligand has decreased to zero. Due to the fast redistribution of ligands, the fraction of ligand bound in the isolated way increases during the dissociation process, thereby increasing the effective dis-

sociation rate. The different shapes of the curves in Fig. 2.6 and Fig. 2.7 can be used to qualify a dissociation process upon observing the dissociation time curves.

The model given here corresponds to the IELE (Instanteneous Establishment of Lattice Equilibrium) approximation given by Epstein [93], as far as only the dissociation is considered. The IELE model itself gives a general time-function of the binding ratio when both association and dissociation take place.

CHAPTER 3

# SPECIFICITY OF THE BINDING OF BACTERIOPHAGE M13 ENCODED GENE- 5 PROTEIN TO DNA AND RNA STUDIED BY MEANS OF FLUORESCENCE TITRATIONS

## 3.1 ABSTRACT

The fluorescence quenching of the bacteriophage M13 encoded gene-5 protein was used to study its binding characteristics to different polynucleotides. Experiments were performed at different salt concentrations and in some instances at different temperatures. The affinity of the protein depends on the base and sugar composition of the polynucleotides involved and may differ appreciably, i.e. by orders of magnitude. The salt dependence of binding is within experimental accuracy equal for all single stranded polynucleotides. A method is presented to estimate values of the cooperativity constant from salt titration curves. These values are systematically higher than those obtained from titration experiments in which protein is added to a polynucleotide solution. A comparison is made between the binding constants of the gene-5 protein and the gene-32 protein encoded by the T4 phage. Possible implications of the binding characteristics of the gene-5 protein for an understanding of its role in vivo are discussed.

## 3.2 INTRODUCTION

To a large extent cellular processes are regulated by protein nucleic acid interactions, but thus far a basic understanding of the underlying mechanisms is rather limited. This situation may change rapidly because crystal structures of proteins involved in protein nucleic acid interactions have been published [41,97-100] or are soon to appear. Also theoretical models for the thermodynamic and kinetic description of their association behaviour have been

proposed [71,89,90,101] and have been applied in a number of experimental studies [35,37,59,92,102,103].

One of the proteins that has been extensively studied is the gene-5 protein encoded by the filamentous phage M13, which uses the E. coli bacterium as a host. During the life cycle of the phage the protein induces a switch from replicative form replication to single stranded DNA synthesis [17,18,24]. It is generally assumed that in the process the protein coats the nascent plus strand to prevent minus strand synthesis. Recently, a second and interesting function of the gene-5 protein has been uncovered. The protein regulates the cellular level of the M13 gene-2 product by means of translational repression [47,48].

It has been a long held belief, that the gene-5 protein is a nonspecific single stranded DNA binding protein. However, recent experiments indicate that this point of view is too restricted. By means of fluorescence binding experiments it has been shown that the gene-5 protein binds stronger to poly(dT) than to poly(dA) by at least two orders of magnitude [37] It has also become apparent the protein is able to bind to double stranded RNA [104]. Moreover, a series of elegant experiments by Fullford and Model suggest that the translational repressor activities of the gene-5 protein may be brought about by sequence specific gene-5 protein RNA interaction [83].

To investigate whether specific binding under physiological conditions may find its origin in a preference of the gene-5 protein for specific bases and/or the sugar composition of the nucleotides we have carried out an extensive series of fluorescence binding experiments. The binding of the gene-5 protein to various synthetic and naturally occurring polynucleotides was studied by means of fluorescence titrations measurements [37] and by means of salt titrations experiments and may be appreciably different for different polynucleotides. A method was developed which allowed us to derive the binding characteristics from the latter type of experiments. Interestingly, the binding preferences of the gene-5 protein differ from those of the analogous double helix destabilizing gene-32 protein from the T4 phage, which was studied in detail by Von Hippel and his collaborators [102,105].

### 3.3a Materials

Gene-5 protein was isolated as described by Garssen et al. [60]. The protein concentration was determined from the absorbance at 276 nm using a molar absorption coefficient of 7100 $M^{-1}$ $cm^{-1}$ [29]. Fluorescence titrations were performed with solutions containing 10 mM sodium cacodylate (pH 6.9) and various concentrations of NaCl. Poly(dA), poly(rA), poly(rU), poly(rC) and poly(rI) were purchased from Boehringer; poly(dC), poly(dU), poly(dI) and poly(dT) from P.L. Biochemicals. M13-DNA was isolated from M13 phages as described by Konings et al. [106]. Yeast $tRNA^{Phe}$ and 5S rRNA were isolated according to standard methods [107-111].

The concentrations of the polynucleotides were determined from U.V. absorption measurements using the following extinction coefficients: poly(dA), 9300 $M^{-1}cm^{-1}$ (260 nm); poly(rA), 9500 $M^{-1}cm^{-1}$ (260 nm); poly(dC), 7200 $M^{-1}cm^{-1}$ (270 nm); poly(rC), 6500 $M^{-1}cm^{-1}$ (267 nm); poly(dU), 9200 $M^{-1}cm^{-1}$ (260 nm); poly(rU), 9200 $M^{-1}cm^{-1}$ (260 nm); poly(dI), 9400 $M^{-1}cm^{-1}$ (260 nm); poly(rI), 9400 $M^{-1}cm^{-1}$ (260 nm); poly(dT), 8700 $M^{-1}cm^{-1}$ (265 nm); M13 DNA, 7400 $M^{-1}cm^{-1}$ (259 nm); yeast $tRNA^{Phe}$, 6800 $M^{-1}cm^{-1}$ (260 nm); 5S rRNA, 6800 $M^{-1}cm^{-1}$ (260 nm).

### 3.3b Fluorescence titrations

Fluorescence measurements were carried out with a Perkin Elmer fluorescence spectrophotometer MPF-4 equipped with a 150 W Xenonlamp. The instrument was used in the so-called ratio mode; this means that relative fluorescence signals were measured instead of absolute quantum yields. Excitation was at the absorption maximum (276 nm) of the tyrosyl residues and the fluorescence was measured at the emission maximum (303 nm). All measurements were performed at 5 $^{o}C$, unless mentioned otherwise. Binding experiments of the gene-5 protein to poly(dI) and poly(rI) were performed after possibly exist-

ing aggregates [112,113] had been removed by heating the solution of the polynucleotide above the presumed melting temperature (15 min. at 70 °C) prior to the titration experiment.

### 3.3c Evaluation of titration curves

Two types of titration experiments were carried out:
1) A solution of polynucleotide was titrated with a protein solution both at the desired NaCl concentration.
2) Mixtures of gene-5 protein and polynucleotides at low ionic strength, containing the polynucleotide in (slight) excess, were titrated with a concentrated NaCl solution. This results in the dissociation of the protein polynucleotide complex, giving rise to an enhancement of the fluorescence intensity of the solution.

The analysis of the binding experiments obtained in the first type of experiments has been described earlier [37]. The fluorescence intensity as a function of the protein concentration is given by

$$F_{obs} = ((1+(\nu[N_t]/[P_f])(1-Q_{max})) / (1+\nu[N_t]/[P_f])) [P_t] F_f \qquad (3.1)$$

where $F_{obs}$ is the observed fluorescence intensity; $F_f$ is the fluorescence intensity per mol of free protein monomers; $[N_t]$ is the total nucleotide concentration expressed in mononucleotides; $[P_t]$ and $[P_f]$ are the total and the free protein concentration (expressed in monomers) respectively; $Q_{max}$ is the fraction of protein fluorescence quenched upon binding; $\nu$ is the degree of saturation per mononucleotide (also called binding density) defined as

$$\nu = [P_b] / [N_t] \qquad (3.2)$$

$[P_b]$ is the concentration of bound protein (monomer). The fluorescence titration experiments are interpreted using the model of McGhee and Von Hippel [71] (see section 2.1a).

$$\nu / [P_f] = K_{int} B \tag{3.3}$$

where $K_{int}$ is the intrinsic binding constant for the binding of a protein to an isolated site on the DNA. In Eq. 3.3 the parameter B is

$$B = (1 - n\nu) \{((2\omega - 1)(1 - n\nu) + \nu - R) / (2(\omega - 1)(1 - n\nu))\}^{n-1}$$

$$\times \{(1 - (n + 1)\nu + R) / 2(1 - n\nu)\}^2 \tag{3.4}$$

The meaning of the symbols is as follows: n is the number of nucleotides covered by one protein monomer, $\omega$ is the cooperativity factor, i.e. the factor by which $K_{int}$ is multiplied if a protein molecule binds adjacent to an already occupied site.

$$R = \{(1 - (n + 1)\nu)^2 + 4\omega\nu(1 - n\nu)\}^{1/2} \tag{3.5}$$

Combination of Eqns. 3.1, 3 2 and 3.3 allows the calculation of the fluorescence intensity as a function of the parameters $K_{int}$, $\omega$, n, $F_f$ and $Q_{max}$. Subsequently, the calculated fluorescence intensity is fitted to the observed intensity determined in experiments as mentioned under a (in order to obtain the best set of values for these parameters).

For the interpretation of the salt titration curve mentioned under b we first consider Eq. 3.3 and 3.4. For sufficiently high values of $\omega$ (i.e. $\omega > n$), at half saturation the effective binding constant is equal to the reciprocal of the free protein concentration [71], i.e.

$$(K_{int}\omega)_{1/2} = [P_{f,\ 1/2}]^{-1} \qquad (3.6)$$

where the subscript indicates the situation of half saturation of the polynucleotide. Hence, by adjusting the salt concentration to obtain the situation of half saturation a binding constant can be derived.

In the further analysis of the salt titration curves (mentioned under b) use is made of the following equation

$$\log(K_{int}\omega) = c \log [M^+] + b \qquad (3.7)$$

Studies of the binding of oligopeptides or proteins to RNA and DNA have shown that the effective binding constant conforms to Eq. 3.7 for salt concentrations in the range 0.1 to 1 M NaCl [114-116]. In Eq. 3.7 $[M^+]$ is the concentration of monovalent cations (in our case $Na^+$) for which the salt dependence of the effective binding constant, $K_{int}\omega$, is studied. Within the $Na^+$ concentration range considered, c and b are constants. Theoretical arguments for the validity of Eq. 3.7 have been provided by Record and collaborators [73,74] (see Discussion). From Eq. 3.7 it follows that at half saturation the cation concentration $[M^+]_{1/2}$ and the corresponding effective binding constant $(K_{int}\omega)_{1/2}$ are related to the effective binding constant $(K_{int}\omega)_1$ at the salt concentration $[M^+]_1$ by

$$(K_{int}\omega)_1 = (K_{int}\omega)_{1/2} ([M^+]_1 / [M^+]_{1/2})^c \qquad (3.8)$$

In our approach c and $(K_{int}\omega)_{1/2}$ are experimentally available parameters. Hence, for a given value of $\omega$ (assuming that $\omega$ is constant throughout the salt titration) $K_{int,1}$ can be calculated at a particular salt concentration $[M^+]_1$. At this salt concentration

$$[P_{f,i}] = [P_t] - \nu_i [N_t] \qquad (3.9)$$

where use has been made of Eq. 3.2. Substitution in Eq. 3.3 yields

$$B_i K_{int,i} ([P_t] - \nu_i [N_t]) - \nu_i = 0 \qquad (3.10)$$

In this equation the only unknown is $\nu_i$, which can be calculated by standard zero point methods. Making use of Eq. 3.13 (see below) and the relation

$$\nu = 0 [P_t] / [N_t] \qquad (3.11)$$

where 0 is the fraction of protein bound to the polynucleotide, the fluorescence intensity at the salt concentration $[M^+]_i$ can be calculated according to

$$F_{calc,i} = F_f ([P_t] - \nu_i [N_t] Q_{max}) \qquad (3.12)$$

Hence starting from Eq. 3.8 the fluorescence as a function of the salt concentration can be calculated for a specific value of $\omega$. By varying $\omega$ the calculated curve can be fitted to the experimental curve, which permits an estimation of $\omega$.

## 3.4 RESULTS

### 3.4a Titration experiments

Titrations of gene-5 protein polynucleotide mixtures with NaCl were carried out for various ribo- and deoxyribohomopolynucleotides and for some natural polynucleotides in order to study the influence of the type of sugar and/or

base on the strength of the protein-nucleotide complexes. At low ionic strength the protein is bound to the polynucleotide; upon increasing the NaCl concentration the gene-5 protein is dissociated from the polynucleotide as witnessed by the increase of the fluorescence intensity. Therefore, in all experiments the conditions were such that the polynucleotide was present in (slight) excess, ensuring that before addition of NaCl all of the protein was bound. In most cases the titration curves level off upon increasing NaCl concentration and the fluorescence intensity approaches a constant value (Figs. 3.1 and 3.2). From the observed fluorescence intensity the fraction of protein bound to polynucleotide, θ, can be calculated with

$$\theta = ([P_t]F_f - F_{obs}) / [P_t]F_fQ_{max} \qquad (3.13)$$

where $[P_t]F_f$ is the fluorescence intensity at the end of the titration and $[P_t]F_fQ_{max}$ is the difference in fluorescence intensity between the beginning and the end of the titration curve.

Fig. 3.3 shows θ for various complexes as a function of the NaCl activity. This figure clearly demonstrates the effect of the type of polynucleotide on the stability of the protein-polynucleotide complex, which is about equal for poly(rC), poly(rA) and poly(dA). The stability of these complexes is lower than for example the stability of the complex with poly(rU). To dissociate the poly(dU) complex still higher salt concentrations are required. The complexes with poly(dC) and particularly with poly(dT) are so stable that even at 2 M NaCl a constant fluorescence level cannot be obtained, indicating that at this concentration the complex is still only partly dissociated (see Fig. 3.2). From the curves in Fig. 3.3 the salt concentrations were deduced at which the polynucleotide lattices are half saturated (see Table 3.1). For poly(dC) and poly(dT) only lower limits can be given since complexes of the protein with these polynucleotides cannot be completely dissociated. All the experiments represented in Figs. 3.1 and 3.2 were performed at a total protein concentration of 10 μM which means that at half saturation $K_{int}{}^{\omega} = 0.2 \times 10^6\ M^{-1}$.

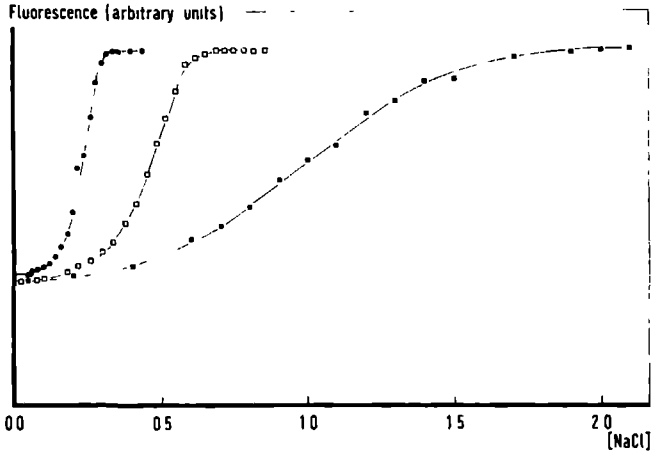



Figure 3.1. Gene-5 protein fluorescence as a function of salt concentration, recorded at 5 °C, upon disruption from poly(dA) (●), poly(rU) (□) or poly(dU) (■).

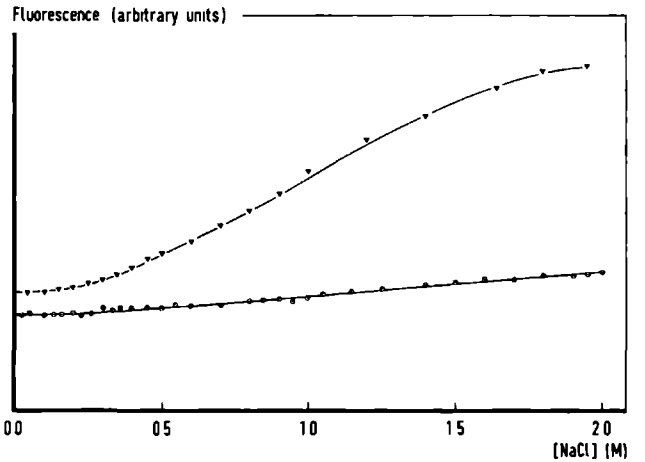



Figure 3.2. Gene-5 protein fluorescence as a function of salt concentration recorded at 5 °C, upon disruption from poly(dC) (▼) or poly(dT) (●).

Values of $K_{int}^{\omega}$ at other salt concentrations were found by changing the total protein concentration (and consequently the total nucleotide concentration) and also by measuring and analysing binding isotherms as described earlier [37]. The salt concentration region most suitable for measuring binding isotherms was also deduced from the titrations with NaCl; this is in fact the region where the dissociation of the complexes takes place. At different salt

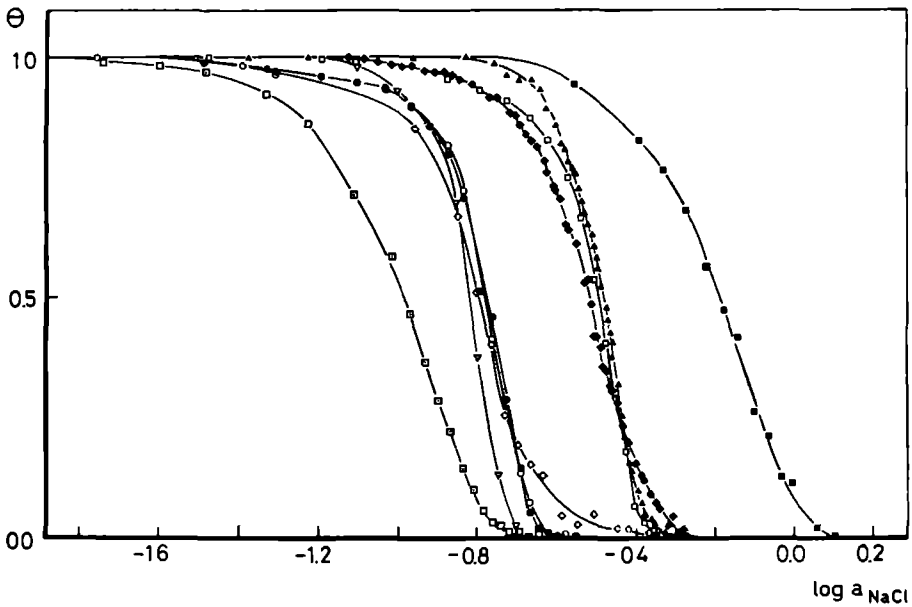



Figure 3.3. Fraction, θ, of gene-5 protein bound to different polynucleotides as a function of the NaCl activity, determined at 5 °C. Total protein concentration of 10 μM poly(rA) (○); poly(dA) (●); poly(rU) (□); poly(dU) (■); poly(dI) (▲); poly(rC) (▽); yeast $tRNA^{Phe}$ (⊡); 5S rRNA (◇) or M13 DNA (◆).

Table 3 I

Concentration of NaCl at half saturation of different polynucleotides at a total gene-5 protein concentration of 10 μM obtained at 5 °C.

| polynucleotide | [NaCl] (M) |
|---|---|
| dA | 0 23 |
| dC | > 1 0 |
| dI | 0 61 |
| dT | > 2 |
| dU | 0 96 |
| rA | 0 23 |
| rC | 0 22 |
| rI | 0 90 |
| rU | 0 47 |
| M13 DNA | 0 44 |
| 5S rRNA | 0 22 |
| yeast $tRNA^{Phe}$ | 0 15 |

concentrations in these regions, titrations with gene-5 protein were performed for the different polynucleotides (not shown). Again the results were analysed with the binding model of McGhee and Von Hippel [71], as described by Alma et al. [37].

For all polynucleotides studied the $Q_{max}$, the fraction of the fluorescence quenched upon binding, is 0.60 ± 0.04, except for poly(dT) where $Q_{max}$ = 0.80. The number of nucleotides, n, covered by one protein molecule turned out to be 4.0 ± 0.3, concurrent with our earlier results. These parameters as well as $K_{int}\omega$ are related to well defined characteristics of the titration curves and can therefore be determined with more accuracy than the cooperativity parameter, ω, which varies around 200 and than $K_{int}$. Therefore the effective binding constant $K_{int}\omega$ was analysed as a function of the salt concentration.

The values obtained for $K_{int}^{\omega}$ as a function of salt concentration are presented in Fig. 3 4, where log $K_{int}^{\omega}$ is plotted against log [NaCl]. Within experimental accuracy linear plots are obtained (sets of two datapoints were left out of consideration). Moreover, again within experimental accuracy, the slopes of these lines are equal (Table 3.II), i.e.

$$d \log (K_{int}^{\omega}) / d \log [NaCl] = -4.0 \pm 0.8 \qquad (3.14)$$

Hence, specificity of the binding of the gene-5 protein to the various polynucleotides is not found in the slopes but in the relative displacement of the lines in Fig. 3.4. The parameter log $K^{o}$, i.e. the value of log ($K_{int}^{\omega}$) at 1 M NaCl, can serve as a quantity to express its specificity, the data are collected in Table 3.II. Note that d log $K_{int}^{\omega}$ / d log [NaCl] and log $K^{o}$ correspond to the constants c and b respectively in Eq. 3.7.

Table 3 II

Binding characteristics for the salt dependence at 5 °C of the binding of gene-5 protein to different polynucleotides (see Eq 3 7)

| polynucleotide | d log ($K_{int}^{\omega}$)/d log [NaCl] = c | log $K^{o}$ = b |
|---|---|---|
| dA | -4 7 | 1 9 |
| dI | -3 3 | 4 4 |
| dU | -4 6 | 5 2 |
| rA | -4 6 | 2 2 |
| rC | -3 5 | 2 7 |
| rI | -3 6 | 4 1 |
| rU | -3 5 | 4 0 |
| M13 DNA | -3 3 | 3 8 |
| yeast $tRNA^{Phe}$ | -3 2 | 2 5 |

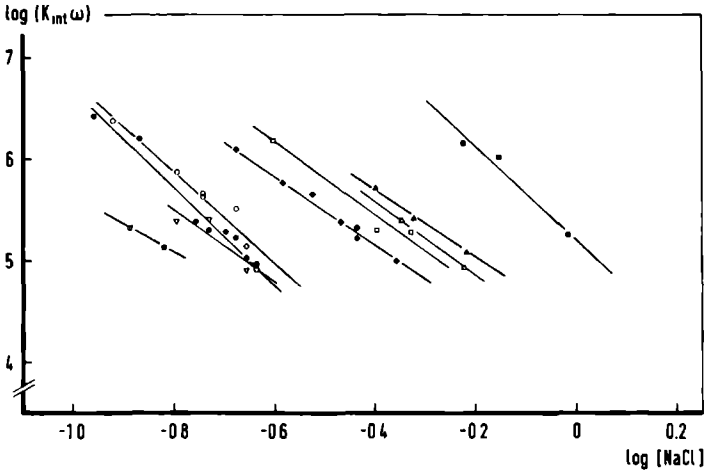



Figure 3.4. Salt dependency of the effective binding constant $K_{int}\omega$ for gene-5 protein polynucleotide complexes, as derived from salt titrations of the complex (e.g. Figure 3.3) and from binding isotherms, at 5 °C. poly(rA) (o); poly(dA) (●); poly(rU) (□); poly(dU) (■); poly(rI) (Δ); poly(dI) (▲); poly(rC) (▽); yeast $tRNA^{Phe}$ (▣); 5S rRNA (◇) and M13 DNA (◆).

### 3.4b Derivation of the cooperativity factor from salt titrations

In the preceding section the salt titration experiments were employed to derive values for the effective binding constant, $K_{int}\omega$, as a function of the salt concentration. As indicated under Materials and Methods (section 3.3c) these curves can also be employed to obtain independent estimates of $K_{int}$ and $\omega$ using the linear dependence of log $K_{int}\omega$ on log [NaCl]. This procedure is necessarily restricted to homopolynucleotides. For heteropolynucleotide lattices it is not possible to define a unique intrinsic binding constant. Its value will vary along the macromolecular chain which precludes a proper simulation of the salt titration curve. A similar remark applies for $\omega$.

As an example, the salt titration curve of the gene-5 protein poly(dA) complex is considered. The fluorescence intensity of the poly(dA) gene-5 protein mixture was calculated as a function of the salt concentration. In the simulations the following constants were introduced: the number of nucleotides covered by one protein monomer n = 4; the salt dependence of the effective binding constant c = -4. The reason for this choice is evident (see preceding section): for all experiments we find that n and -c are equal to four within experimental accuracy. The effect of different values of $\omega$ on the (simulated) salt titration curves is shown in Fig. 3.5. Excellent correspondence between experimental and simulated curve is obtained for $\omega \approx 500$.

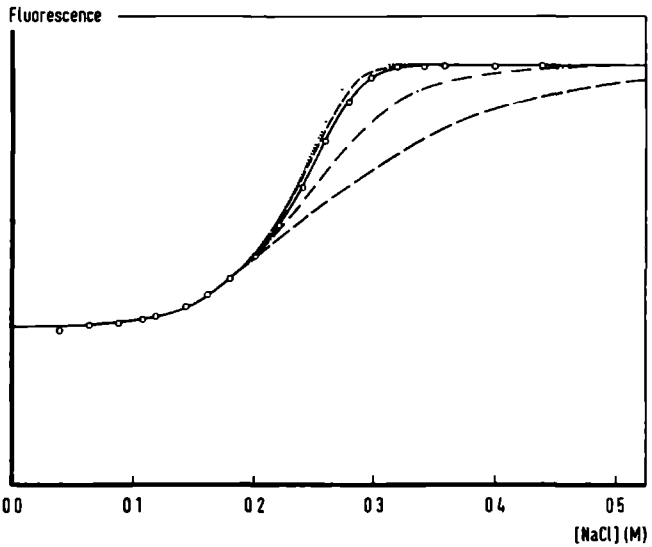



Figure 3.5. Simulation of salt titration curves for different values of $\omega$; $\omega$ = 5 (- - -); 50 (-·-·-); 500 (—); 1500 (-----) and 5000 (.....). (o) represent experimental points obtained at 5 °C for the gene-5 protein poly(dA) complex.

Examination of Fig. 3.5 shows that the first part of the titration curve, i.e. at low salt concentrations, is rather insensitive to the value of ω. However, at higher salt concentrations, i.e. when the degree of saturation becomes less than 0 5, the form of the curves is much more affected by ω. This behaviour can be explained qualitatively as follows. Titration of the protein polynucleotide complex with a concentrated salt solution results in the breakage of protein-nucleotide contacts mainly at the end of a protein cluster, provided that ω is large enough to prevent the disruption of the binding of a doubly contiguous bound protein. During the titration the degree of saturation decreases which is followed by a redistribution of the bound protein to more and therefore smaller clusters However, this redistribution to smaller clusters will be inhibited for increasing values of ω. This means that for progressively higher values of ω steeper salt titration curves will be observed at the end of the titration, increasingly approaching an all or none behaviour. The theoretical salt titration curves of the other gene-5 protein homopolynucleotide complexes show similar behaviour as is discussed above for the gene-5 protein poly(dA) system. The cooperativity factors deduced from the salt titrations are listed in Table 3 III together with those obtained by analysing the convential binding isotherms.

Table 3 III

Gene-5 protein cooperativity factors obtained from salt titrations and binding isotherms which show the best correspondence of the simulated and experimental curves In all salt titration simulations c is taken -4 0

| polynucleotide | salt titrations | binding isotherms (average values) |
|---|---|---|
| dA | 500 | 130 |
| rA | 770 | 110 |
| rU | 220 | 120 |
| rC | --- | 100 |

### 3.4c Temperature dependency of the complex formation

Former experiments [37] showed that the enthalpy effect for the binding of the gene-5 protein to poly(dA) is close to zero. In contrast, we find that for the gene-5 protein poly(dU) complex formation studied at 0.70 M NaCl and at different temperatures (see Table 3.IV) $\Delta H$ = -17.2 kcal/mol, as calculated from a Van 't Hoff plot. At 25 °C three salt titrations were performed with gene-5 protein M13 DNA complexes. The measured binding constants are virtually the same as the binding constants obtained at 5 °C from interpolation of the data in Table 3.II, indicating very small enthalpy effects upon the binding of the gene-5 protein to its viral genome (see Table 3.IV)

Table 3 IV

Thermodynamic parameters characterizing the binding of the gene-5 protein to various polynucleotides

| polynucleotide | [NaCl] (M) | $K_{int}^{\omega}$ ($10^5$ $M^{-1}$) | $\Delta H^o$ (kcal $mol^{-1}$) | $\Delta S^o$ (cal $mol^{-1}$ $K^{-1}$) |
|---|---|---|---|---|
| dA | 0 175 | 3 5 (25 °C) | + 2 | 32 |
| dA | 0 21 | 1 3 (25 °C) | - 2 | 17 |
| dU | 0 70 | 11 1 ( 5 °C) | -17 2 | -34 |
| dU | 0 70 | 3 21 (15 °C) | -17 2 | -34 |
| dU | 0 70 | 1 32 (25 °C) | -17 2 | -34 |
| M13 DNA | 0 28 | 4 95 (25 °C) | + 0 5 | 27 |
| M13 DNA | 0 33 | 2 53 (25 °C) | - 0 4 | 23 |
| M13 DNA | 0 42 | 0 87 (25 °C) | - 2 2 | 15 |

## 3.5 DISCUSSION

### 3.5a Binding characteristics

The present paper is an extension of our earlier fluorescence binding experiments with the aim of obtaining a more detailed insight into the binding characteristics of the gene-5 protein. Within experimental accuracy we find from all our experiments that the number of nucleotides covered by one protein monomer is 4. This is in agreement with our earlier results and with those of a number of other laboratories [17,29,33,35,59,63]. In cases where higher binding ratios are found the lattice is (most likely) not completely occupied. At this time it is not clear how the n = 4 binding ratio can be matched with the detailed structural model for the gene-5 protein DNA interaction advocated by Brayer and McPherson which suggest a binding ratio of n = 5 [41].

One aspect that has not been recognized when applying the McGhee and Von Hippel theory in the analysis of binding isotherms of polynucleotide protein complexes is the polyelectrolyte nature of the polynucleotides. The present binding experiments and those performed in other laboratories provide ample evidence that the binding of gene-5 protein is to a large extent determined by electrostatic interactions. Electrostatic interactions will be much more effective at the start of a titration experiment, when the DNA molecule is virtually unoccupied than at the end of the titration when the DNA lattice is almost saturated. Consequently, upon binding of the protein the apparent binding constant gradually diminishes because the electrostatic contribution to the binding becomes less. This introduces an anticooperativity in the binding of the protein to the DNA, a phenomenon that is well known for for instance $Mg^{2+}$ binding to tRNA [117,118]. This anticooperative binding effect will moderate the cooperativity of the protein binding, resulting in lower (apparent) values for $\omega$ than obtained through the salt titration curves.

In a proper description of the complex formation of the gene-5 protein with polynucleotides the release of counterions from the polynucleotide (and possibly anions from the protein) should be accounted for. Therefore the complex formation reaction is represented by

$$\text{protein} + \text{polynucleotide} \rightleftarrows \text{complex} + (m'\psi + k)\ \text{ions} \qquad (3.15)$$

$m'$ is the number of ion pairs formed per protein molecule and $\psi$ is the average number of cations bound per phosphate group. Consequently, $m'\psi$ is the number of monovalent cations released from the polynucleotide; $k$ is the number of monovalent anions released from the protein. According to Record and coworkers [74,115]

$$\log (K_{int}{}^{\omega}) = -(m'\psi + k) \log [M^+] + \log K^o \qquad (3.16)$$

where $[M^+]$ is the monovalent cation concentration. We found that within experimental accuracy $c = -(m'\psi + k) = -4$. If we assume that $m' = n = 4$ and make use of the fact that $\psi = 0.75$ [73] then the number of sodium ions released upon binding one protein molecule is 3, while the protein looses one $Cl^-$-ion; this applies for all polynucleotides. Calculations of the first term of r.h.s. of Eq. 3.16 (for instance at 0.2 M $Na^+$) also shows that the electrostatic contribution to the binding is appreciable.

### 3.5b Specificity of the binding

The results presented in this paper provide ample evidence for the existence of large variations in the affinity of the gene-5 protein for different nucleotide sequences. Some rules can be formulated to describe these variations in binding strength. It is pointed out from the start that the simple proposition that binding of the protein to DNA is stronger than to RNA does not obtain in general. Examination of Fig. 3.4 demonstrates that the gene-5 protein has a higher affinity for poly(rU) and poly(rI) than for poly(dA). However, it does follow from the present experiments that the binding to polyribonucleotides is weaker (or equal) than to the polydeoxyribonucleotide homologues, i.e. the order in binding strength is: poly(dA) = poly(rA); poly(dC) > poly(rC); poly(dU) > poly(rU); poly(dI) > poly(rI).

The observed differences in binding strength can only be explained to a limited extent. The weakest binding observed is that to tRNA. This is reasonable because tRNA has a strong tendency to form double helices and tertiary structure so that particularly at somewhat higher salt concentrations gene-5 protein binding cannot compete with structure formation in tRNA.

The relatively weak binding to poly(dA) and poly(rA) compared to for instance binding to poly(dU) has been explained before. The polyadenylic acids have a strong tendency to form highly stacked structures. In the nucleoprotein complex these stacking interactions are disrupted and most likely replaced by stacking between the adenine bases and aromatic amino acid residues of the protein. This followed from NMR experiments but is also suggested by the observation that the enthalpy change upon binding of the gene-5 protein is close to zero. This is interpreted to mean that the enthalpy loss occurring upon destacking of the adenine bases is compensated for by the enthalpy gain resulting from the stacking of these bases on the aromatic amino acids In concurrence with this interpretation it is observed that binding to poly(dU) is accompanied by an enthalpy change of $\Delta H$ = -17.2 kcal/mol. At ambient salt concentrations the bases in poly(dU) are not stacked so that the favourable enthalpy change caused by the possible stacking of the aromatic amino acids on the uridines may become manifest. As yet we are not able to offer an explanation for other differences in the affinity of the protein towards the various polynucleotides. Thus it is not clear as to why the protein binds to a particular polydeoxyribonucleotide in preference to the corresponding polyribonucleotide. Also the much higher affinity for poly(dT) than for poly(dU) cannot be accounted for. The latter characteristic is not only a property exhibited by the gene-5 protein. Similar observations have been made for the helix destabilizing gene-32 protein encoded by bacteriophage T4 [102] as well as for the E. coli single strand binding protein [119].

Thymine differs from uracyl only in that a methyl group is attached to the C5 ring atom instead of a proton and yet this small difference results in a gene-5 protein poly(dT) complex which cannot be dissociated at salt concentrations at which the gene-5 protein poly(dU) complex is fully disrupted. The hydrophobic nature of the methyl group and a more effective stacking of the thymidine in the complex could be responsible for this increase in binding strength

[120]. It is interesting to note that quenching of the gene-5 and gene-32 protein fluorescence is higher upon binding to poly(dT) than to other polynucleotides.

Another interesting feature concerning the deplacement of the protein from DNA follows from an examination of Fig. 3.3. The salt titration curve acquired for the M13 DNA protein complex is less steep than those of the two polynucleotides, i.e. poly(rU) and poly(dI), with similar effective binding constants. We consider this to be a demonstration of the heterogeneity in the base composition of the M13 DNA. During the salt titration first the weaker binding nucleotide residues (e.g. stretches of A-residues) will loose the bound gene-5 protein, resulting in a decrease of the degree of saturation. With increasing salt concentrations this process continues so that the fraction of the nucleotides to which the gene-5 protein shows stronger binding gradually increases. In the end this leads to the tailing in the salt titration curves at low degrees of saturation.

### 3.5c Comparison with the gene-32 protein

Just as the gene-5 protein the gene-32 protein encoded by the T4 phage is a double helix destabilizing protein. Its DNA and RNA binding properties have been studied in even more detail than those of the gene-5 protein and it is interesting to make a comparison between the properties of these two proteins.

Just as for the gene-5 protein it was found that binding of the gene-32 protein to deoxyribopolynucleotides is stronger than to their ribopolynucleotide homologues. However, in contrast to the gene-5 protein the affinity of the gene-32 protein for DNA is always higher than for RNA polynucleotides. As we have seen in the preceding section both proteins bind much more effectively to poly(dT) than to poly(dU). Effective binding constants obtained for gene-32 and gene-5 protein to various polynucleotides at 0.23 M NaCl and 37 $^{o}$C are presented in Table 3.V. The available data show that the affinity of the gene-5 protein to the various nucleotides may differ significantly from that of the gene-32 protein

Table 3 V

Comparison of the binding constants $K_{int}\omega$ ($M^{-1}$) of gene-5 protein and gene-32 protein [102,105] polynucleotide complexes at 37 °C and 0 23 M NaCl The data represent extrapolated values The temperature dependency for the binding constant of the gene-5 protein to polyribonucleotides was assumed to be the same as to the corresponding polydeoxyribonucleotides For calculations discussed in the text (see Fig 3 8) cooperativity values of $\omega$ = 500 for gene-5 protein and $\omega$ = 2000 for the gene-32 protein are used

| polynucleotide | gene-5 protein | gene-32 protein |
|---|---|---|
| dA | $9\ 2\ 10^4$ | $2\ 10^7$ |
| dC | --- | $2\ 10^8$ |
| dU | $6\ 0\ 10^6$ | $5\ 10^7$ |
| rA | $1\ 3\ 10^5$ | $3\ 10^6$ |
| rC | $1\ 0\ 10^5$ (*) | $3\ 10^4$ |
| rU | $8\ 0\ 10^4$ | $4\ 10^6$ |
| M13 DNA | $8\ 7\ 10^5$ | --- |
| T4 DNA | --- | $8\ 10^7$ |

*) holds only if $\Delta H = 0$

### 3.5d Possible implications for the in vivo behaviour of the gene-5 protein

The quantitative binding parameters now available for the gene-5 protein permit us to reflect on some aspects of the role of the gene-5 protein during replication and translation. The present experiments demonstrate that the protein can bind to DNA as well as to RNA and that it is even capable of "unwinding" the tRNA molecule. Indifferent binding to RNA and DNA could be lethal to the cell, but our data indicate that the molecule is so beautifully designed that, at the ionic conditions and gene-5 protein concentration prevailing in the cell, this does not seem a serious problem. This is illustrated in Fig. 3.6 for the binding of the protein to tRNA. The effective binding

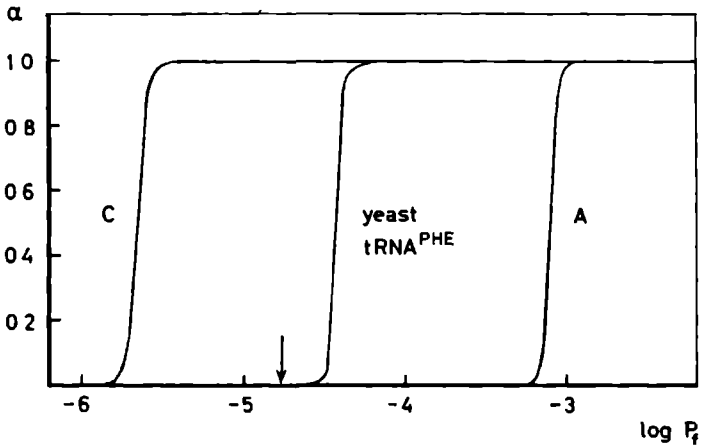



Figure 3 6. Simulation of the degree of saturation, α, of hairpin A (A), hairpin C (C) (see text) and yeast $tRNA^{Phe}$ as functions of the free gene-5 protein concentration, $[P_f]$, at physiological conditions (37 °C, 0 23 M NaCl) The arrow indicates physiological concentrations of the gene-5 protein when single stranded DNA synthesis is occurring

constant for the complex formation of gene-5 protein to yeast $tRNA^{Phe}$ at 0.2 M NaCl can be estimated from Fig. 3.4. This value was used to calculate the degree of saturation of the tRNA with the protein as a function of the free protein concentration. The calculation shows that the protein concentration has to be increased beyond the estimated free protein concentration in the cell (17 μM) [37] before binding to tRNA can take place. In practice the binding isotherm will be shifted more to the right because in the cell tRNA will predominantly be complexed to proteins, e.g. elongation factor Tu (EF-Tu) or aminoacyl tRNA synthetases.

Similar considerations can be extended to the hairpins which can be formed by the viral genome. A number of these stem-loop structures are surmised to act as regulatory signals e.g. a hairpin structure downstream of gene-8 func-

tions as a rho-independent transcription termination signal [121,122]. At this time we confine our attention to hairpins which may be formed by the complementary sequences in the intergenic region between gene-4 and 2. Experimental evidence for the existence of these hairpins, which are presented in Fig. 3.7, is given in the literature [123-127]. In particular we consider the influence of the gene-5 protein on hairpins A and C. The fraction of hairpins, $\alpha$, melted out by the protein is given by

$$\alpha = [P_f]^m K \; / \; ([P_f]^m K + 1) \tag{3.17}$$

where m is the maximum number of protein molecules that can bind to a nucleotide chain of the length of the hairpin. K is the equilibrium constant characterizing the reaction

$$H + mP \rightleftarrows H_s P_m \tag{3.18}$$

where H stands for the hairpin and $H_s P_m$ for the melted hairpin complexed to m protein molecules. To calculate K we have adopted the following model: the hairpin is converted into the single stranded state and subsequently completely occupied by gene-5 protein. Then

$$K = K_1 K_2 \tag{3.19}$$

where $K_1$ is the equilibrium constant characterizing the disruption of the hairpin and $K_2$ is the equilibrium constant characterizing the complete saturation of the disrupted hairpin with gene-5 protein.

$$K_2 = S \; K_{int} \; (K_{int}{}^{\omega})^{m-1} \tag{3.20}$$

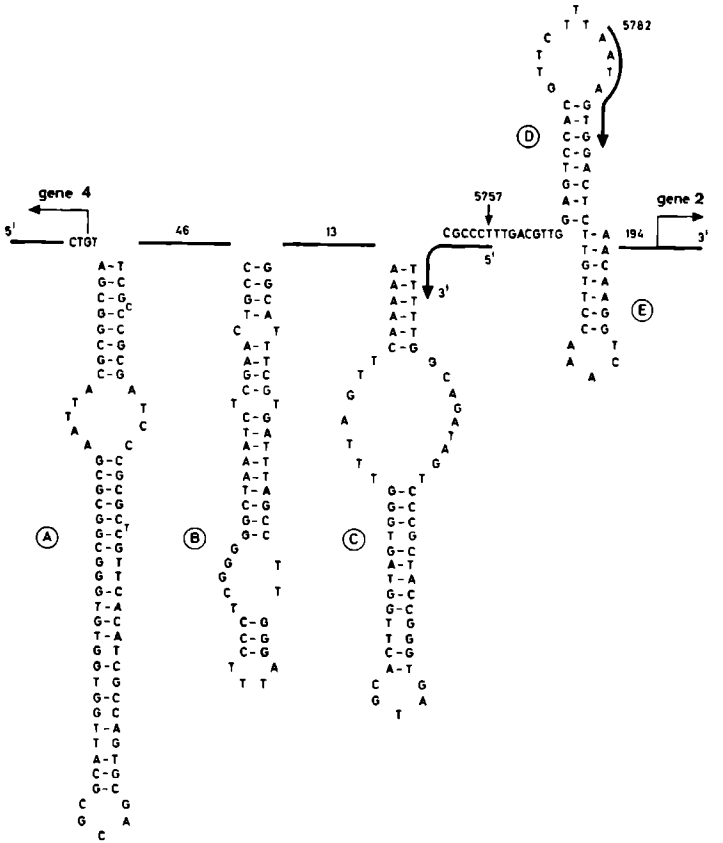



Figure 3.7. The intergenic region of the M13 genome with its proposed secondary structures (hairpins A to E). Positions 5757 and 5782 are the starting points of complementary strand respectively viral strand synthesis; denoted by bold arrows. Single stranded parts containing no secondary structure are represented by a straight line; the number of nucleotides composing these structures is indicated.

S is a statistical factor giving the number of ways one protein cluster can bind to the polynucleotide. Hence $K_2$ can be calculated from the average effective binding constant obtained for M13 DNA at 0.23 M NaCl and 37 °C (Table 3.V). $K_1$ can be estimated from the thermodynamic parameters available for the formation of DNA hairpins [128-130]

In Fig. 3.6 $\alpha$ is given for hairpin A and C (Fig. 3.7) as a function of the free protein concentration. The protein concentration required to melt hairpin A far exceeds the free protein concentration in the cell. This indicates that this hairpin is not melted out in the gene-5 protein M13 DNA complex. Hence, the gene-5 protein does not seem to impair the possible role of this hairpin during phage assembly that has been suggested by many authors [2,12,131].

On the other hand hairpin C is melted out at the concentrations of gene- 5 protein prevailing in the cell after DNA synthesis is switched from replicative form replication to single stranded DNA synthesis. Therefore at these gene-5 protein concentrations the hairpin probably can no longer serve as the origin for RNA primer synthesis needed for replicative form replication [124].

In addition to its role during phage DNA replication the gene-5 protein has been implicated to act as a translational repressor for the gene-2 protein. Analogous properties are exhibited by the gene-32 protein encoded by the T4 phage [126,132-134]. As has been indicated in the previous section the gene-32 protein is a helix destabilizing protein which plays a role in DNA replication. It also acts as a translational repressor and thereby regulates its own synthesis. Von Hippel et al. have offered an explanation for this behaviour, not by invoking the existence of specific binding sites which derive their specificity from complementary matrices of hydrogen bond donors and acceptors, but by utilizing the binding cooperativity to amplify modest differences in binding affinities to different types of nucleic acids and different sequences [105]. As a result the protein can bind preferentially to a particular unstructured region around the ribosome binding site of its mRNA.

Based on the thesis that the gene-5 protein behaves very similarly to the gene-32 protein, Fulford and Model investigated whether the proteins i.e. the

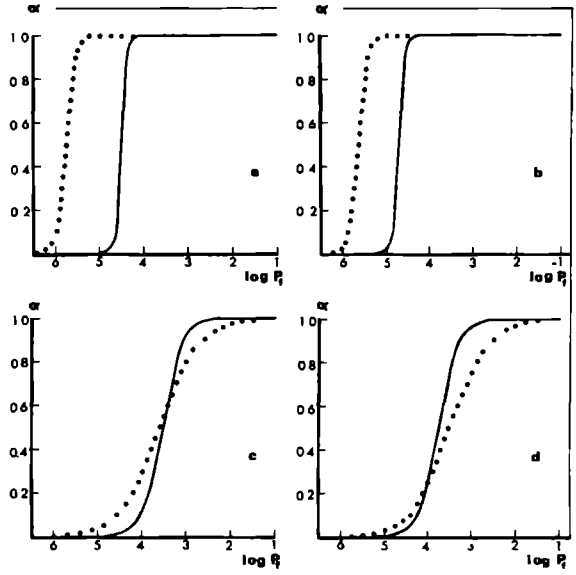


Figure 3 8 Simulation of the degree of saturation, α, of homo-oligonucleotides
a) oligo-$(rU)_{28}$
b) oligo-$(rA)_{28}$,
c) oligo-$(rU)_8$,
d) oligo-$(rA)_8$
as a function of the free protein concentration $[P_f]$ Saturation by gene-32 protein is indicated by and saturation by gene-5 protein by — For the calculatation the binding constants listed in Table 3 V were used

gene-32 protein and the gene-5 protein could replace each other in their translational repressor functions [83]. The answer turned out to be negative which led to the suggestion that preferential binding to specific RNA sequences (or structures) could play a role in the repression activities The results obtained in this paper for the gene-5 protein may cast some doubt on the underlying assumption. In some instances the binding properties of the gene-5 protein differ significantly from those of the gene-32 protein (see Table 3.V). Particularly noteworthy is our observation that at 5 °C (see Fig. 3.4) the affinity of the gene-5 protein to poly(rU) (and poly(rI)) is slightly higher than the average affinity to M13 DNA and much higher than to poly(dA) and poly(rA). For the gene-32 protein the order of affinities is quite different (vide supra). Guided by the proposal of Fulford and Model concerning the regulatory RNA sequences for the gene-32 and gene-2 mRNAs, we considered two simplified and extreme situations We examined the binding of the gene-5 and the gene-32 protein to a stretch oligo(rU) and oligo(rA) consisting of 28 and of 8 nucleotides, which we considered as models for the regulatory mRNA sequences. The results are presented in Fig. 3 8 For the

28 nucleotide long molecules the gene-32 protein first fills up the lattice; in both cases the concentration of gene-5 protein required to achieve the same degree of saturation is much higher. For the short oligonucleotides $(rA)_8$ and $(rU)_8$ the differences in concentrations of the two proteins at which saturation is attained are too small to explain the specificity observed in the translation repression experiments. Hence, by this approach we cannot mimic the results of Fulford and Model in both examples, which may be taken to support their thesis that the proteins may bind preferentially to a certain sequence (or structure) of the mRNA.

# DIFFERENT DNA BINDING MODES AND COOPERATIVITIES FOR BACTERIOPHAGE M13 GENE-5 PROTEIN REVEALED BY MEANS OF FLUORESCENCE DEPOLARISATION STUDIES

## 4.1 ABSTRACT

The binding of the bacteriophage-M13-encoded gene-5 protein to oligo(deoxythymidylic acid)s and M13 DNA was studied by means of tyrosyl fluorescence decay and fluorescence anisotropy measurements. The observed fluorescence decays could be described with two exponentials, characterised by the lifetimes $\tau_1$ = 2.2 ns and $\tau_2$ = 0.8 ns respectively. Only the amplitude of the longer-lifetime component is influenced by binding of the protein to DNA. This indicates that a part of the tyrosyl residues is involved in the binding. By means of fluorescence depolarisation measurements the rotational correlation time of the protein dimer is found to be 12.9 ns. In contrast to earlier measurements carried out on the DNA binding protein of phage Pf1 by Kneale and Wijnaendts van Resandt [84], the observed rotational correlation times of the gene-5 protein pass through a maximum when the protein is titrated with oligo(deoxythymidylic acid)s. This is not observed upon titration with M13 DNA. Our measurements showed that for the oligo(deoxythymidylic acid)s there clearly is a decrease in the number of clustered proteins on the lattice in the case of excess nucleotide. This is a direct consequence of the much lower cooperativity of the binding to the oligonucleotides compared to the cooperativity characteristic of binding to polynucleotides. The number of nucleotides covered by a protein monomer is found to be $\leq 3$ for the oligonucleotides and $\approx 4$ for M13 DNA. Model calculations show that the "time-window" through which the fluorescence depolarisation can be observed (i.e. the fluorescence lifetime) in this case significantly affects the "measured" effective rotational correlation times.

## 4.2 INTRODUCTION

An interesting aspect of the interaction between the single strand DNA binding gene-5 protein, encoded by the bacteriophage M13, is its ability to bind to DNA in two different modes, previously designated the oligonucleotide and the polynucleotide binding mode. In the polynucleotide binding mode the protein covers four nucleotides and the cooperativity of binding is high, i.e. the binding of a protein adjacent to an already bound protein is increased by a factor of five hundred ($\omega$ = 500) with respect to binding to a naked lattice. In the so called oligonucleotide binding mode the protein covers three nucleotides and the cooperativity of the binding is two orders of magnitude smaller ($\omega$ = 5). Moreover, the salt dependence of the two binding modes differs significantly [37,67,Chapter 3 of this thesis].

Recently the binding characteristics of the single strand DNA binding protein of the filamentous phage Pf1 were studied in detail by means of fluorescence depolarisation measurements [84,135]. Similarly to the M13 gene-5 protein and the T4 single strand DNA binding gene-32 protein [102,103,136], the Pf1 protein exhibits highly cooperative binding to single strand DNA. However, in contrast to the M13 gene-5 protein, for the Pf1 protein different binding modes could not be distinghuished: complex formation to polynucleotides and oligonucleotides turned out to be very much the same. Thus the question arises whether the two proteins are intrinsically different in their DNA binding properties. This is not self evident, because the second binding mode discovered for the M13 gene-5 protein was detected in a series of NMR measurements on the complex formation with oligo(deoxyadenylic acid)s [67]. For this method much higher concentrations ($10^{-3}$ M) of nucleic acid and protein are required than for fluorescence measurements ($10^{-6}$ M) and it was suggested either that binding properties may be concentration dependent, so that the possible second binding mode for the Pf1 protein could not be picked up in the fluorescence experiments, or that the shifts observed in the NMR spectra upon complex formation between the gene-5 protein and oligo(deoxyadenylic acid)s might result from protein-protein interaction instead of protein-DNA interaction [84]. In our opinion the Overhauser effects observed between protein and nucleotide residues [69] argue against the latter interpretation. To obtain further insight into the importance of the suggestions mentioned

above the interaction between gene-5 protein from the M13 phage and DNA was studied by fluorescence depolarisation as well. The results of this study are presented in the present paper and they confirm the interpretation of the earlier NMR experiments [67,69]. An interesting and new aspect of these fluorescence depolarisation studies is that the decrease of the number of proteins clustered on the DNA lattice can be observed directly. Moreover the gene-5 protein is one of the first examples of a protein with only tyrosines as fluorescent probes which is studied by means of fluorescence depolarisation. Tyrosyl residues have a relatively short fluorescence lifetime. In general the fluorescence lifetime influences the value of the rotational correlation time calculated from fluorescence depolarisation measurements. This is particularly manifest for short fluorescence lifetimes and this effect will be demonstrated in the present study.

## 4.3 MATERIALS AND METHODS

### 4.3a Materials

Gene-5 protein was isolated and purified as described previously [60], except that during Escherichia coli sonication no DNAse I was added. Concentrations of the gene-5 protein were measured using a molar extinction coefficient at the absorption maximum (276 nm) of 7100 $M^{-1}$ $cm^{-1}$. The oligo(dexoythymidylic acid)s were purchased from P-L Biochemicals. They have 5'-terminal phosphates. The extinction coefficient for $(dT)_8$ at 266 nm is $\varepsilon$ = 8690 $M^{-1}$ $cm^{-1}$ [137]. The extinction coefficients $\varepsilon$ = 8600 $M^{-1}$ $cm^{-1}$ for $(dT)_{12}$ and $\varepsilon$ = 8540 $M^{-1}$ $cm^{-1}$ for $(dT)_{18}$ at 266 nm were obtained from interpolation of literature data [137]. M13 DNA was purified according to Konings et al. [107]; a molar extinction coefficient of 7400 $M^{-1}$ $cm^{-1}$ at 259 nm was used for this molecule.

All measurements were performed in a buffer containing 10 mM sodium cacodylate, 50 mM sodium chloride (unless otherwise mentioned) at pH 6.9.

### 4.3b Fluorescence measurements

Fluorescence lifetimes and anisotropies were measured at 20 °C with the EMBL fluorescence spectrometer, which is described in detail elsewhere [138]. The system consists of a synchronously pumped cavity dumped dye laser, the output of which is frequency doubled to 285-310 nm. The pulse width of the system is 10 ps, with a repetition rate of 4 MHz. The tyrosine fluorescence of the sample was excited at 285 nm and observed through a 20 nm wide interference filter centered around 300 nm.

Fluorescence decay curves were measured using the single photon technique [139] with a very fast channelplate photomultiplier (Hamamatou R1564U). The time response of the system is 70 ps (full width half maximum). Since the time response of the system is very short (less than one channel in the data) the model function was not convoluted with the apparatus response.

The vertically and horizontally polarised components of the emitted light were detected simultaneously by the same detector. This is done by using a 45 ns optical delay of the horizontally polarised component with respect to the vertically polarised emitted light. The time dependent change of the fluorescence anisotropy, A(t), and of the fluorescence itself, F(t), upon excitation with a very short, plane-polarised pulse, is given by (see e.g. [140,141])

$$A(t) = (I_{//}(t) - I_{\perp}(t)) \; / \; (I_{//}(t) + 2\, I_{\perp}(t)) \qquad (4.1)$$

and

$$F(t) = I_{//}(t) + 2\, I_{\perp}(t) \qquad (4.2)$$

where $I_{//}$ is the intensity of the light polarised parallel and $I_{\perp}$ the intensity polarised perpendicularly to the plane of polarisation. The data were analysed using a non-linear least squares fitting procedure, SPLMOD by R. Vogel [142].

## 4.4 RESULTS

### 4.4a Fluorescence lifetime measurements

**An example of a fluorescence intensity decay curve measured for the M13 gene-5 protein is shown in Fig. 4.1a. All observed fluorescence decay curves**

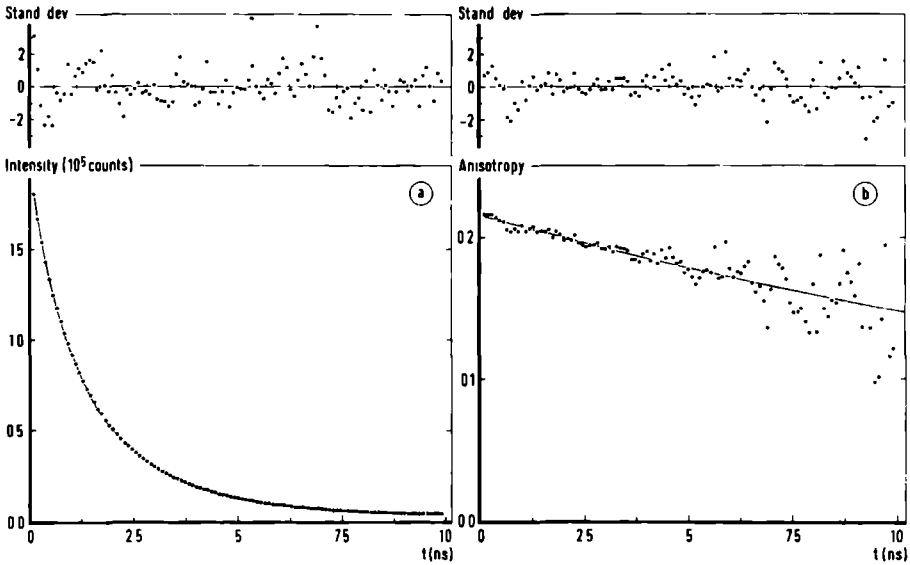



Figure 4.1. a) Example of a fluorescence decay curve of the gene-5 protein; the excitation wavelength was 285 nm and the emission wavelength was 300 nm. b) Example of a fluorescence anisotropy decay curve of the gene-5 protein; excitation and emission wavelengths as under (a). Both curves were fitted by means of a least squares analysis program [142] for sums of exponentials. The weighted residuals of these fits are shown above the curves. For the curve in (a) the results were $\tau_1$ = 2.10 ± 0.02 ns, $A_1$ = 0.62 ± 0.01, $\tau_2$ = 0.80 ± 0.02 ns, $A_2$ = 0.38 ± 0.01 with a reduced standard deviation of fits, $\chi^2$ = 1.03 (see Eq. 4.3). For the curve in (b) we obtained $\phi$ = 26.4 ± 1.0 ns with a reduced standard deviation of fit, $\chi^2$ = 0.98.

could be adequately described by a linear combination of two exponentials. It was found that the two decay constants of these exponentials were independent of the size and composition of the protein-DNA complex and that only the relative amplitudes of the components were affected. Hence, all fluorescence decay data could be described by

$$F(t) = A_1 e^{-t/\tau_1} + A_2 e^{-t/\tau_2} \qquad (4.3)$$

with $\tau_1 = 2.2 \pm 0.2$ ns and $\tau_2 = 0.8 \pm 0.1$ ns. It was found that the amplitude of the short-lived component was independent of the degree of saturation of nucleotide with gene-5 protein, while the amplitude of the long-lived component depended upon the degree of binding. This is illustrated more clearly in Fig. 4.2 for the titration of the gene-5 protein with $(dT)_{12}$

It is noted that for steady state measurements the contribution of the fast component is very small compared with that of the slow component since the total fluorescence is proportional to the surface of the entire decay curves. For example, for the free protein dimer the contribution of the fast component is 15 % of the total time average fluorescence. This contribution increases in the case of the protein-polynucleotide complex as a result of the quenching of the slow component. Further analysis and possible interpretations of the change in fluorescence decay curves upon protein complex formation are outside the scope of this chapter.

### 4.4b Fluorescence depolarisation measurements

An example of the decay of the fluorescence anisotropy of the gene-5 protein is presented in Fig. 4 1b. The time dependence of the fluorescence depolarisation signal of the free protein can be described by [141]

$$A(t) = A(0) e^{-t/\phi} \qquad (4.4)$$

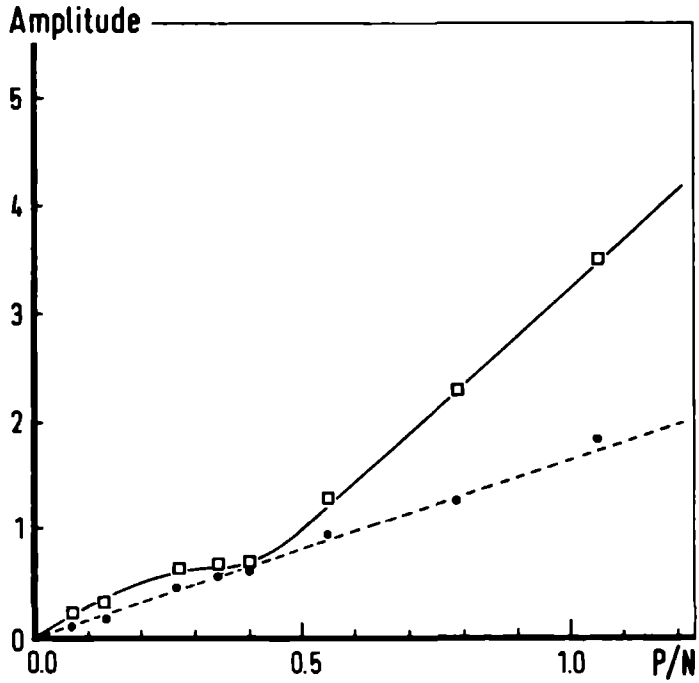



Figure 4.2. Amplitudes $A_1$ (□) and $A_2$ (•) (see Eq. 4.3) of the slow ($\tau_1$ = 2.2 ns) and fast ($\tau_2$ = 0.8 ns) fluorescence decays of the protein as a function of the ratio of the concentration of (mono)nucleotides and protein monomer obtained upon titration of $(dT)_{12}$ with gene-5 protein. The $(dT)_{12}$ concentration was kept constant at 67.93 μM throughout the titration. The decay curves were analysed according to Eq. 4.3.

**where A(0) is the zero time anisotropy and $\phi$ is the rotational correlation time of the protein. For globular particles**

$$\phi = f_{rot} V_h \eta / kT \qquad (4.5)$$

where $V_h$ is the hydrated particle volume, $\eta$ is the viscosity of the sample and $f_{rot}$ is the relative rotational friction coefficient, which is a measure for the deviation from sphericity. For spherical molecules $f_{rot} = 1$. The hydrated specific volume of a typical protein is about 1 $cm^3\ g^{-1}$, therefore $V_h$ = M / $(6x10^{23})$ $cm^3$. For dilute protein samples one can take the viscosity of water at the temperature used (1.002 cpoise at 20 °C). The theoretical rotational correlation time for a spherical protein with the same mass as the gene-5 protein dimer (M = 19378 g $mol^{-1}$) then is $\phi$(20 °C) = 8.00 ns. Actual measurements of the fluorescence decay of the gene-5 protein (see Fig. 4.1b) yield $\phi$ = 12.9 ns at 79 µM and 15.4 ns at 840 µM protein concentration. At the low concentration this results in a rotational friction coefficient of 1.61. The higher value of 15.4 ns at 840 µM is interpreted to arise from protein aggregation.

The effect of salt concentration on protein aggregation was also studied. Only a small increase of $\phi$ was found. At a protein concentration of 39 µM $\phi$ increased from 12.9 ns at 50 mM NaCl to 14.8 ns at 1.9 M NaCl. The increase of sample viscosity will be the main cause for this effect. At a protein concentration of 840 µM these values were 15.4 ns and 24.5 ns respectively. We conclude that at these protein concentrations a considerable amount of protein tetramers will be present in solution at 1.9 M NaCl.

### 4.4c Titration experiments

The effect of complex formation between the gene-5 protein and $(dT)_8$, $(dT)_{12}$, $(dT)_{18}$ or M13 DNA on the rotational correlation time was studied by measuring the fluorescence depolarisation as a function of the nucleotide protein ratio, N/P, at constant nucleotide concentration. Each individual fluorescence titration point was measured with a freshly prepared sample to avoid artifacts due to protein irradiation damage. Within experimental error the time dependence of the fluorescence anisotropy could be fitted with a single exponential. The fluorescence lifetime of tyrosine residues is relatively short (see Fig. 4.1a),thus a relatively small "window" is available for measur-

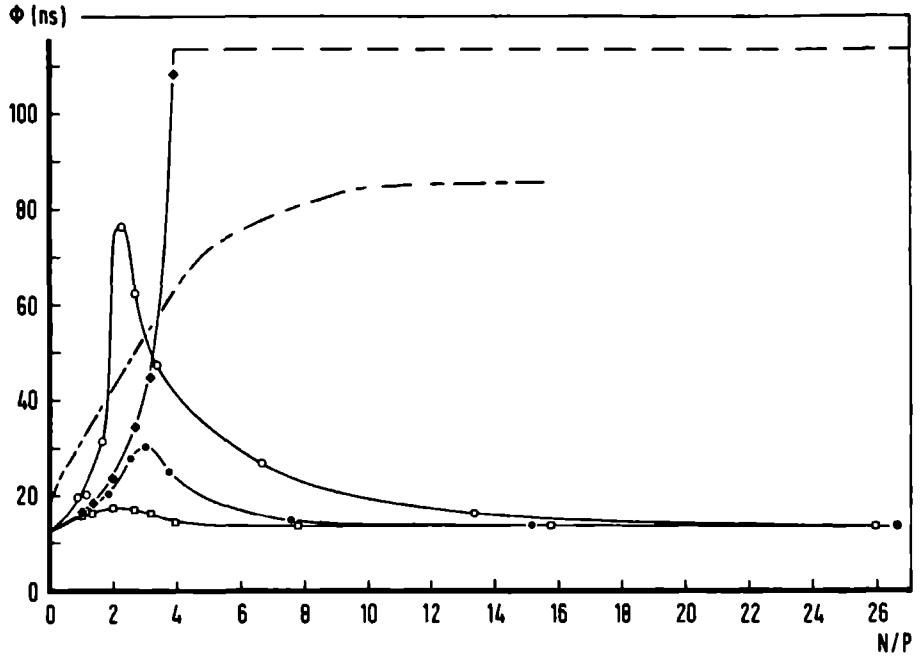



Figure 4 3 Rotational correlation times, $\phi$, of gene-5 protein/nucleotide mixtures as a function of N/P, the ratio mononucleotide/protein monomer, obtained at 20 °C The nucleotide concentration (expressed in mononucleotides) was kept constant during the titration at 70 43 µM $(dT)_8$ (□), 67 93 µM $(dT)_{12}$ (•), 59 72 µM $(dT)_{18}$ (o) or 70 47 µM M13 DNA (◆) The broken line for the M13 DNA curve at higher N/P values indicates high values of $\phi$ The exact values cannot be determined because of the fast decaying protein fluorescence For comparison also results are given for the binding of the Pf1 protein to $(dT)_{16}$ (adapted from [84]) at a constant protein concentration of 18 9 µM (– - – -)

ing the fluorescence depolarisation. In practice this means that, for the experiments described in this paper, only rotational correlation times smaller

than about 50 ns can be measured with reasonable accuracy. During the titration experiments protein complexes with different molecular weights are present in solution. This means that the measured single exponential is characterised by an effective correlation time. The dependence of this correlation time upon the N/P ratio is presented in Fig. 4.3. The $\phi$ values of the complexes formed between the gene-5 protein and $(dT)_8$, $(dT)_{12}$ or $(dT)_{18}$ pass through a maximum when the N/P ratio is increased. Subsequently, at higher N/P values $\phi$ gradually decreases to 13.7 ns for all oligonucleotide complexes studied, which is a slightly higher value than that obtained for the free protein. This is in contrast to earlier experience with the Pf1 DNA-binding protein [84], for which $\phi$ first increases and then reaches a constant value for the higher N/P ratios. For reasons of comparison the rotational correlation time obtained upon titration of the Pf1-DNA binding protein with $(dT)_{16}$ is added in Fig. 4.3. Upon titration of the M13 gene-5 protein with M13 DNA large $\phi$ values are obtained at higher N/P ratios (indicated by the dashed line in Fig. 4.3). Because of the small window inherent to tyrosyl fluorescence depolarisation, $\phi$ values could not be quantified in this situation.

In one case, namely for the complex formation of the gene-5 protein and $(dT)_{18}$, the reverse experiment, i.e. a titration where nucleotide is added to a protein solution at constant protein concentration, was performed. Within experimental error the same results were obtained for the N/P dependence of the depolarisation decay as for the titration experiments described above.

## 4.5 DISCUSSION

The most surprising finding of the present experiments is that the effective rotational correlation times obtained for the oligonucleotide gene-5 protein complexes pass through a maximum for increasing N/P ratios and return to a value close to that obtained for the free protein. This is in contrast to the results obtained for the gene-5 protein M13 DNA complexes, for which $\phi$ does not decrease to the free protein value at high N/P ratios and is also in contrast to the results obtained for the Pf1 DNA binding protein complexes formed with oligo- and polynucleotides. A qualitative and straightforward explanation for these observations is as follows. For low N/P ratios the protein

is present in excess and the effective correlation time corresponds to that of the free protein. When the N/P ratio is raised, the concentrations of the gene-5 protein DNA complexes increase, i.e. particles with a longer rotational correlation time contribute to an increase of the effective correlation time. This process continues until, roughly speaking, the equivalence point of about three (oligo(dT)) or four (M13 DNA) mononucleotides per protein molecule are present in solution. Then for increasing N/P values, i.e. for excess nucleotide concentrations, a redistribution of the protein molecules among the DNA strands takes place when the cooperativity factor is not of sufficient magnitude. Subsequently, one protein dimer per DNA strand will be found for high N/P values. This is the situation for the gene-5 protein oligo(dT) complex formation. For high values of the cooperativity factor a redistribution of the DNA binding protein among the different DNA lattices will not take place or will do so to a much lesser extent Hence, to a first approximation only DNA strands fully occupied by protein and naked DNA strands will be present in solution. This situation is exemplified by the experiments conducted for the Pf1 DNA binding protein (Fig. 4.3 and [84]).

To obtain a better insight into the role of the cooperativity and the effect of the fluorescence lifetime on the measured effective correlation time, model calculations were performed for the complex formation between a single strand DNA binding protein and a nonanucleotide. It was assumed that the intrinsic binding constant of the protein molecule to the oligonucleotide amounts to $10^6$ $M^{-1}$ and that the number of nucleotides covered by one protein molecule equals three. The calculations were performed for a total protein concentration of 100 μM.

In general, the fluorescence depolarisation for a (slowly) exchanging system in equilibrium is given by

$$A(t) = \Sigma_i \; A_i(0) \; f_i \; e^{-t/\phi_i} \qquad (4\,6)$$

where $A_i(0)$ is the zero time anisotropy, $\phi_i$ is the rotational correlation time of complex i and $f_i$ is the fraction of the protein present in complex i. During

the titration process, complexes of the type SP, $SP_2$ and $SP_3$ can be formed between the nonanucleotide strand, S, and the protein, P. Therefore, for the case at hand i may vary from 0 (i.e. free protein) to 3. For practical purposes Eq. 4.6 can be simplified to

$$A(t) = A(0) \Sigma_i f_i e^{-t/\phi_i} \qquad (4.7)$$

because it turned out (not shown) that for the gene-5 protein/oligo(dT) systems the zero-time anisotropy varied to a limited amount (maximum total decrease was only 14 % throughout the titration). Theoretical curves for the rotational correlation time, $\phi$, as a function of the N/P ratio can now be simulated as follows. First, A(t) (Eq. 4.7) is calculated with the following input parameters. The fraction, $f_i$, of the protein present in the particle of type i, is calculated from a combination of the law of mass conservation and expressions for the binding constants obtained from Epstein's theory [87] for the binding of a ligand to an oligonucleotide as a function of oligonucleotide concentration. The rotational correlation time $\phi$ was assumed to be 10 ns, 12 ns, 22 ns and 32 ns for i = 0 to 3, values which are thought to be close to those expected for complexes between the gene-5 protein and a nonanucleotide. Combination of these different parameters allows a calculation of the anisotropy A(t) as a function of time, using Eq. 4.7, and as a function of nucleotide concentration. Subsequently the effective rotational correlation time is calculated by fitting a mono-exponential function to the curve calculated for A(t). The values thus found for the effective rotational correlation time as a function of the nucleotide protein ratio are plotted in Fig. 4.4; three different situations are presented. Curve (a) represents the results of a calculation in which the cooperativity factor was assumed to be unity (i.e. no cooperativity). For this situation a square-window parameter of 7 ns was used. The window parameter is equal to the time interval in which the mono-exponential is fitted to the decay curve. In practice this is determined by the fluorescence lifetime. It is seen that the effective rotational correlation time goes through a maximum as expected. When this situation (absence of cooperativity is compared with the situation of high cooperativity ($\omega$ = 500), with the same window of 7 ns, curve (b) is obtained. Again the effective rotational

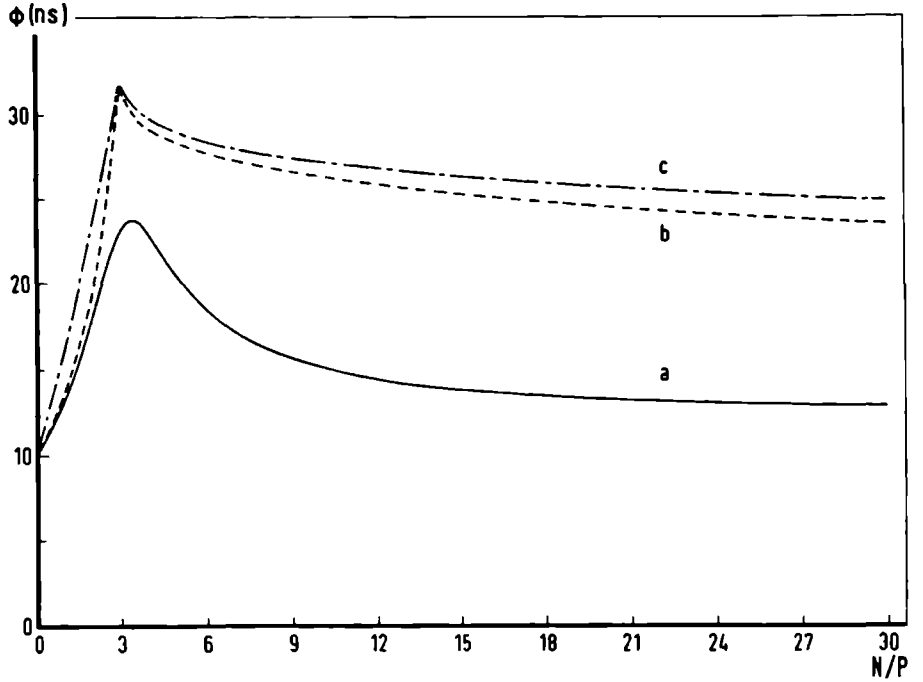



Figure 4.4. Rotational correlation times calculated for a single strand DNA binding protein as a function of N/P, the ratio mononucleotide/protein molecule, for binding to nonanucleotides. It was asssumed that one protein molecule covers three nucleotides when bound to the DNA. The rotational correlation times employed in the calculation are 10, 12, 22 and 32 ns for the free protein and the complexes SP, $SP_2$ and $SP_3$ respectively, where S denotes the nonanucleotide strand and P the protein. The protein concentration was kept constant at 100 μM. An intrinsic binding constant of $10^6$ $M^{-1}$ and different values of the cooperativity parameter were used in the calculations. The φ values were obtained through a one-exponential fit to the simulated anisotropy decay curves for (a) a cooperativity constant ω = 1 and window = 7 ns, (b) ω = 500 and window = 7 ns, and (c) ω = 500 and window = 60 ns (see text).

correlation time passes through a maximum but does not decrease to the low values of $\phi$ observed in absence of cooperativity for high values of N/P. The sharp maximum is higher than in absence of cooperativity and is located exactly at N/P = 3, corresponding to the binding stoichiometry of the system. Finally, we consider again a situation with $\omega$ = 500 but now with a window of 60 ns. This yields curve (c), in which $\phi$ exhibits a sharp maximum located at N/P = 3. At increased values of N/P, $\phi$ diminishes but not as much as for curve (b), this is caused by the increase of the window parameter. Also under the conditions valid for curve (c), the convex downward character of curves (a) and (b) at lower N/P ratios has disappeared. This is the consequence of fitting a mono-exponential function to the actual decay curves (Eq. 4.7). In particular at low N/P values, where the amplitudes of the more slowly decaying components are small, small windows will overemphasize the contribution of the fastest decaying component (in our case the free protein) giving rise to convex downward curves. This situation does not occur when application of larger windows is possible as is demonstrated by curve (c).

It is noted in passing that in view of the assumptions made quantitative agreement between the experimental and the theoretical results is not to be expected. However, the model calculations account for the most important features observed for the effective rotational correlation time of the gene-5 protein and the Pf1 protein (Fig. 4.3). For gene-5 protein binding to the oligonucleotides and M13 DNA convex downward curves are observed at low N/P ratios. This effect can now be attributed to the small window available for the description of anisotropy decay curves, which is brought about by the relatively short fluorescence lifetime of the tyrosine residues. The tryptophan in the Pf1 protein has a longer fluorescence lifetime (7.75 ns for the free protein dimer and 6.52 to 7.42 ns in the complex form [84]) so that a much larger window is available for the description of the anisotropy decay curves. As a result, at low N/P ratios the curve for the Pf1 protein is not convex downward (Fig. 4.3). The cooperativity factor for Pf1 protein DNA binding is ≈ 100 [84] and as expected at higher N/P values no decrease in $\phi$ values is observed (Fig. 4.3). The small maximum observed for curve (c) (Fig. 4.4) is not found for the Pf1 protein. There may be several reasons for this. First, the appearance of the maximum is model dependent in the sense that the form of the maximum strongly depends on the values of the chosen

binding parameters. Secondly, a small maximum as observed in curve (c) (Fig. 4.4) may easily be missed in an actual experiment. Thirdly, curve (c) has been calculated assuming that the $A_i$ values (Eq. 4.6) are more or less equal, which is not applicable to the Pf1 protein [84].

As has been mentioned above at high N/P ratios and high values of the cooperativity parameter, $\omega$, the magnitude of the available window may influence to some extent the calculated $\phi$ values (Fig. 4.4), but we have no experimental material available to check these predictions. For low values of $\omega$ and high N/P ratios the situation is dramatically different. The $\phi$ values decrease significantly and from the model calculations we may conclude that for the gene-5 protein oligo(deoxythymidine) complexes the binding cooperativity (corresponding with a binding mode of n = 2 to 3 (vide infra)) has decreased by about two orders of magnitude. Another interesting feature observed in the calculated curves in Fig. 4.4 is the position of the maxima, which fall at a ratio N/P = 3. This is just the number of nucleotides assumed in the calculations to be covered by the protein. In particular for high values of $\omega$ and small window parameters the $\phi$ curve (starting from low N/P ratios) asymptotically approaches N/P values corresponding to the binding stoichiometry (Fig. 4.4). Returning to Fig. 4.3 we see that the $\phi$ curve obtained for the titration of M13 DNA with gene-5 protein approaches a maximum at N/P = 4. According to the theoretical calculations discussed above this indicates that the binding stoichiometry for this complex is four, in very nice agreement with the results of our earlier fluorescence titration experiments [37,Chapter 3 of this thesis] and experiments of other investigators [17,29,33,35,59,63]. The results obtained for the gene-5 protein oligo(dT) systems are dramatically different. Judged from the location of the maximum in the $\phi$ curves (Fig. 4.3) the number of nucleotides covered by a protein monomer is between 2 and 3.

Although these conclusions are necessarily of a qualitative nature they beautifully confirm the conclusions from our earlier NMR studies on oligo(dA) systems [67,69]. Hence it is now well established that the gene-5 protein may bind in different binding modes, which give rise to different binding cooperativities. The decrease in the number of clustered proteins on the lattice in the case of excess of nucleotides clearly demonstrates that there is a competition between the effect of the cooperativity factor and the number of ways a

set of protein dimers can bind to a DNA lattice (entropy factor). The detailed molecular interpretation of this phenomenon awaits further investigation. In this respect we mention that it has been found recently that even for the binding of the gene-5 protein to polynucleotides under certain circumstances an n = 3 binding mode may occur [143,144]. As an interesting aside it is noted that the fluorescence depolarisation experiments presented in this chapter provide the first direct experimental evidence for a redistribution of protein over the DNA lattice at high lattice concentrations as predicted by the McGhee and Von Hippel theory [71] and by the theory of Epstein [87].

## 4.6 APPENDIX

In the theoretical calculation of the rotational correlation times the fraction $f_i$ of the protein present in complex i served as an input parameter (of Eq. 4.7). Below we indicate the derivation of these parameters. First consider the equilibrium

$$S + iP \underset{\leftarrow}{\overset{K_i}{\rightarrow}} SP_i \qquad (4.8)$$

where P denotes the protein which binds to the nonanucleotide that was used as the lattice in our model calculations. It is assumed that in the complex the protein covers three nucleotides so that i varies from 1 to 3. To be able to calculate the concentrations of the complexes $SP_i$ we introduce the intrinsic binding constant $K_{int}$, which characterises the complex formation of one protein to an isolated binding site on the lattice, and the cooperativity constant $\omega$, which defines the factor by which the binding constant of the protein increases when it binds adjacent to an already bound protein. The binding constants for the formation of the different complexes as defined by Eq. 4.8, can now be found as follows. The complex SP can be formed in seven different ways (the naked lattice has seven binding sites available) so that $K_1 = 7 K_{int}$. For the complex $SP_2$ there are six different ways to bind two isolated proteins and four different ways to bind two proteins adjacent to one an-

other; hence $K_2 = Kint^2 (6 + 4\omega)$. There is only a single $SP_3$ complex possible with $K_3 = Kint^3\omega^2$.
Making use of

$$[S_t] = [S] + \Sigma_i [SP_i] \tag{4.9}$$

$$[P_t] = [P] + \Sigma_i \, i \, [SP_i] \tag{4.10}$$

and

$$[SP_i] = K_i [S] [P]^i \tag{4.11}$$

where $[S_t]$ and $[P_t]$ are the total nonanucleotide and protein concentrations present in the sample, it is easy do demonstrate that

$$[P_t] = [P] + [S_t] ( \Sigma_i \, i \, K_i [P]^i ) / ( 1 + \Sigma_i K_i [P]^i ) \tag{4.12}$$

Given $[P_t]$, $[S_t]$ and $K_i$ (see text), [P] and subsequently [S] and $[SP_i]$ can be calculated which in turn yield the $f_i$ values. The $f_i$ values so obtained were used in Eq. 4.7.

# THE KINETICS OF THE DISSOCIATION OF GENE-5 PROTEIN - POLYNUCLEOTIDE COMPLEXES UPON ADDITION OF SALT

## 5.1 ABSTRACT

The irreversible dissociation kinetics of complexes of M13 encoded gene-5 protein with the polynucleotides poly(dA) and M13 DNA was studied by means of stopped-flow experiments. A linear decay was found for all gene-5 protein - poly(dA) complexes and for the gene-5 protein - M13 DNA complexes for which the lattice was completely saturated at the beginning of the dissociation experiments. Only at the end of the dissociation curve a deviation from linearity was observed. A single-exponential decay was found for the dissociation of gene-5 protein - M13 DNA complexes when the DNA was not completely saturated initially. These results obtained for complexes of gene-5 protein with poly(dA) or with M13 DNA could be interpreted by assuming that dissociation of bound protein is only possible from isolated binding sites, while during this dissociation rearrangement of bound protein clusters continuously takes place, including the formation of newly isolated bound protein. It is not a condition that during this process a lattice equilibrium exists according to the theory of McGhee and Von Hippel [71].

The binding of gene-5 protein to poly(dA) or poly(dT) does not result in a broadening of the nucleotide resonances in the NMR spectra of these polynucleotides, as had been observed for E. coli DNA binding protein and interpreted as an indication for a high rate of translocation of the protein on the polynucleotide [145]. The reason for the absence of line broadening for gene-5 protein - polynucleotide complexes, though the stopped-flow experiments also point to a fast translocation of the protein, is found in the large binding cooperativity in this case. As a consequence of this the major part of

the protein is bound in a cluster which makes the concentration of isolated bound protein very low. This results in a decrease of the signal to noise ratio at higher degrees of binding, but does not lead to line broadening.

## 5.2 INTRODUCTION

For a complete understanding of the molecular details of protein - nucleic acid interactions it is necessary to have kinetic as well as equilibrium data available, since not only the binding strength but also the dynamics of the interaction may be important for its biological function. More specifically, in the case of the gene-5 protein - DNA interaction it is obvious that the thermodynamic characteristics of the binding alone cannot fully explain the role of the protein in the phage DNA replication cycle. For, as a consequence of the high cooperativity of the binding [37, chapter 3 of this thesis] a small change in the gene-5 protein concentration is sufficient to switch between an all or none binding to the viral single strand DNA. Since the binding of the protein inhibits the formation of new replicative form (RF) DNA this would lead to an abrupt termination of the replication process upon increasing protein concentration. However, it is known that the switch from RF replication to single strand DNA synthesis and phage assembling occurs more gradually in vivo [146]. To overcome this difficulty the dynamics of the gene-5 protein - DNA binding has been taken in consideration. According to Shimamoto and Utiyama [59] at an early stage of infection, the host encoded single strand binding (SSB) protein binds to the viral DNA, and RF synthesis takes place. After the formation of a small amount of gene-5 protein some of the SSB proteins on the viral DNA will be replaced by gene-5 proteins and this will result in the formation of a small amount of single strand viral DNA. Later on in infection, when the pool of gene-5 proteins has grown, a complete switch to single strand DNA synthesis occurs. However, this switch can only take place if the supply of gene-5 protein is fast in relation to the rate of growing of the single strand DNA (1000 nucleotides $s^{-1}$) [11]. Because of possible binding site overlap a redistribution of the bound gene-5 proteins is necessary to obtain complete saturation. A mechanism of dissociation of isolated bound proteins followed by an association contiguous to already bound proteins could describe such a redistribution. However, a faster and more efficient redistribution re-

sulting in a fast saturation of DNA with gene-5 protein would be provided by the rearrangement of complete protein clusters [59].

To obtain experimental evidence for such a fast rearrangement, facilitated by translocation (sliding) of the protein along the nucleotide lattice, Pörschke and Rauh [35] and Shimamoto and Utiyama [59] studied the association of gene-5 protein with certain polynucleotides by means of the stopped-flow technique. Because of the different mechanisms involved in the association process, i.e. nucleation, growth and rearrangement coupled with dissociation, it is difficult to explain these results unambiguously. For example, Pörschke and Rauh had to use Monte Carlo simulations to obtain rate constants for the gene-5 protein - poly(dT) and for the gene-5 protein - poly(rA) association reactions. Despite these difficulties the studies mentioned above suggest some fast facilitated mechanism for the redistribution of gene-5 protein.

A different approach for the detection of some fast translocation process has been used by Lohman for the interaction of T4 encoded gene-32 protein with DNA [92,94,96]. He studied the kinetics of the dissociation of protein - DNA complexes with the stopped-flow technique. The dissociation of the complexes was induced by NaCl concentration jumps. Due to the high value of the cooperativity factor (around 5000 for gene-32 protein) it can be expected that only dissociation of protein from a cluster end (single contiguously bound protein) will occur, either directly or via sliding from an isolated bound intermediate. The dissociation of double contiguous molecules from the interior of clusters is too slow to be important while the concentration of isolated protein is negligible. For the gene-32 protein - DNA system this model indeed turned to work out properly: at high NaCl concentration the protein dissociates directly from a cluster end while at low NaCl concentration the dissociation proceeds via the sliding pathway. Therefore we used the same method in order to detect a possible fast rearrangement of gene-5 protein molecules along the DNA lattice.

## 5.3 MATERIALS AND METHODS

### 5.3a Materials

Gene-5 protein was isolated and purified as described previously [60], except that during Escherichia coli sonication no DNAse I was added. Concentrations of the gene-5 protein were measured using a molar absorption coefficient of 7100 $M^{-1}$ $cm^{-1}$ at 276 nm. Poly(dA) and poly(dT) were purchased from P.L. Biochemicals. The concentrations of the polynucleotides were obtained from UV absorption measurements using a molar absorption coefficient for poly(dA) of 9300 $M^{-1}$ $cm^{-1}$ at 260 nm and for poly(dT) of 8700 $M^{-1}$ $cm^{-1}$ at 265 nm. The poly(dA) was at least 1000 nucleotides long as appeared from an analysis on 6 % polyacrylamide gel after phosphate labeling of the polynucleotide. It was relatively homogeneous in length and free from low molecular weight species. M13 DNA was purified according to Konings et al. [106]; a molar absorption coefficient of 7400 $M^{-1}$ $cm^{-1}$ at 259 nm was used for this polynucleotide.

### 5.3b Stopped flow experiments

Stopped flow experiments were performed at 20 $^{o}C$ on a modified Durrum stopped flow apparatus. Equal volumes of two solutions (100 µl each) were rapidly mixed in a cell with an optical pathway of 2 cm and a total volume of 60 µl. The total volume of the mixture (200 µl) was more than the cell volume of 60 µl, so that in the process the cell was made free from remainings of the previous solution. One solution contained 10 mM sodium cacodylate, pH 6.9 and a gene-5 protein - polynucleotide complex in which the degree of saturation of the polynucleotide was 1.00, 0.75 or 0.50. The second solution contained 10 mM sodium cacodylate, pH 6.9 and sodium chloride to a high concentration. The final concentration of the salt after mixing was always high enough to be sure that complete dissociation of the complex occurred. The final concentrations of the polynucleotide were in the range of 100 to 250 µM. For the complexes of the gene-5 protein with poly(dA) the dissociation of the complex resulted in a hypochromic effect on the nucleotide UV absorption at 260 nm, due to the formation of base stacking interactions in the free polynucleotide. Therefore the dissociation experiment could be followed by measur-

ing the UV absorbance. The dissociation process was also followed by monitoring the protein fluorescence since the fluorescence intensity of the free protein is higher than that of the bound protein. Excitation was performed at 276 nm and the fluorescence was measured at the emission maximum of 303 nm. A cut off filter prevented light with a wavelength below 295 nm to enter the detector. Both methods gave exactly the same results. In all experiments the detected signals were recorded and stored into two transient recorders (Biomation) each equipped with a 2 K memory and adjusted to different time ranges to facilitate the simultaneous registration of processes with a large difference in rate. The mixing time, i.e. the time between the start of the mixing process and the first detection, was 3 to 4 ms. Within this period complete mixing was obtained. The data were processed and analysed on a Digital PDP 11-23 computer (Minc). First the signals were scaled between 0 and 1 The curves were smoothed by fitting a seven degree polynomial to the experimental points. From the resulting curves the relaxation times were determined.

The scaled curves obtained from the nucleotide absorbance measurements are identical to the curves measured by means of the protein fluorescence, provided that for the change of the absorbance as well as for the change of the fluorescence intensity a linear relationship exists with the amount of free and bound protein. Then the observed time-dependent absorbance at 260 nm, E(t), equals

$$E(t) = \{\varepsilon_b\,[N_b(t)] + \varepsilon_f\,(\,[N_t] - [N_b(t)]\,)\}.\ell + E_p \qquad (5.1)$$

in which the total nucleotide concentration and the concentration of nucleotide covered with protein are denoted by $[N_t]$ and $[N_b]$. $\ell$ is the optical path length. The absorption coefficients of the nucleotide in the free and in the complexed form are $\varepsilon_f$ and $\varepsilon_b$ respectively and $E_p$ is a constant resulting from the protein absorption at 260 nm. At the end of the dissociation experiment $[N_b(\infty)]$ is zero. The scaled experimental curve is obtained from

$$S(t) = 1 - ( (E(t) - E(0) ) / ( E(\infty) - E(0) ) \quad (5.2)$$

which can be converted using Eq. 5.1 into

$$S(t) = [N_b(t)] / [N_b(0)] = \alpha(t) / \alpha(0) \quad (5.3)$$

in which $\alpha$ is the fraction of the nucleotide lattices covered with protein, i.e. the degree of saturation. An analogous procedure can be followed for the measurements in which the fluorescence is used to monitor the dissociation of the protein - polynucleotide complex. The observed time-dependent protein fluorescence intensity is given by

$$F(t) = F_b [P_b(t)] + F_f ( [P_t] - [P_b(t)] ) \quad (5.4)$$

In this equation the fluorescence intensity per concentration unit of free and bound protein is indicated by $F_f$ and $F_b$, while $[P_t]$ and $[P_b]$ are the concentrations of total and bound protein respectively. The scaled experimental curve is then obtained from

$$S(t) = 1 - ( F(t) - F(0) ) / ( F(\infty) - F(0) ) \quad (5.5)$$

Using Eq. 5.4 in conjunction with the parameter n, which equals the number of nucleotides covered by one protein monomer at complete saturation and with $[P_b(\infty)] = 0$, Eq. 5.5 can be converted into

$$S(t) = [P_b(t)] / [P_b(0)] = [N_b(t)] / [N_b(0)] = \alpha(t) / \alpha(0) \quad (5.6)$$

provided that at t = 0 all the protein present in the solution is bound to the polynucleotide. Thus it appears that after scaling, as described for both the absorption and the fluorescence measurements, the dissociation curves represent the decay of the degree of saturation of the nucleotide.

### 5.3c $^1$H NMR spectra.

Proton NMR spectra were measured on a Bruker CXP-300 spectrometer, interfaced to an Aspect 2000 computer and operating at 300 MHz in the Fourier transform mode. The 90$^o$ pulse length was about 7 μs. For each spectrum about 1000 transients were accumulated using 4K data points. The samples contained 10 mM sodium cacodylate, pH 6.9, and were measured in 99 % $D_2O$. The sample temperature was 22 $^o$C. The spectra were not resolution enhanced.

## 5.4 RESULTS

### 5.4a Stopped flow measurements of the irreversible dissociation of gene-5 protein - poly(dA) and gene-5 protein - M13 DNA complexes

From previous studies performed in our laboratory it has become clear that there are large differences in the binding affinity of gene-5 protein for various polynucleotides [37,Chapter 3 of this thesis]. Among other things this appears from the magnitude of the salt concentration needed to dissociate the protein - polynucleotide complexes at a particular protein concentration. For example, for a total protein concentration of 10 μM the salt concentration at which 50 % of the complex is dissociated amounts to 0.23 M NaCl for poly(dA) and 0 44 M NaCl for M13 DNA. Poly(dA) can thus be considered as a polynucleotide with a relatively low binding affinity for gene-5 protein. Because at a final salt concentration of 0.30 M NaCl or higher almost complete dissociation is obtained for the gene-5 protein - poly(dA) complex, the dissociation process can be studied as a function of the ionic strength in a large salt concentration range, including values approaching the physiological situation. Therefore poly(dA) was chosen as a model for these kinetic studies. Since for the in vivo situation the binding to M13 DNA is important also the dissociation

of gene-5 protein - M13 DNA complexes was followed. In Fig. 5.1A the decay curve for the dissociation of the gene-5 protein - poly(dA) complex is given for an experiment in which the UV absorbance is used for detection, starting at complete saturation. This trace was obtained directly from the data stored in the transient recorder. Fig. 5.1B shows the experimental curve after processing of the data as described in the Materials and Methods (section 5.3b). Through the points a linear decay curve was drawn which represents the first 80 % of the measured decay. The results obtained for the same complex, but now starting at half saturation, are given in Fig. 5.2 in a similar manner. Also in this case the absorbance was used to monitor the dissociation process. Identical curves were observed when the protein fluorescence was followed. The dissociation curve in Fig. 5.2 could be described by a linear decay for about the first 50 % of the time course. These results are charac-

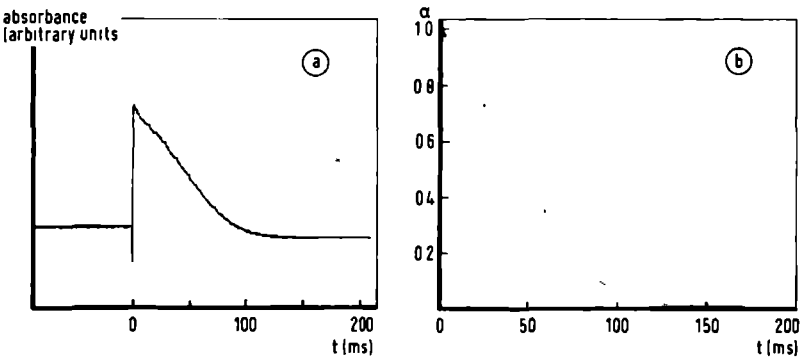



Fig. 5.1. Dissociation curve for the complex of poly(dA) and gene-5 protein at an initial degree of saturation of 1.0. Dissociation is effected by increasing the concentration of NaCl to 2 M. The total nucleotide concentration is 100 μM. A. Experimental decay curves (absorbance versus time). B. Decay of the degree of saturation (α), calculated from (A). The linear curve which fits the largest part of the experimental time course is also given.

teristic for poly(dA) at all conditions at which the experiments were carried out, i.e. at an initial degree of saturation of 1.00, 0.75 or 0.50 and at a final salt concentration of 1.00, 0.75, 0.50, 0.375 or 0.25 M NaCl. The length of the linear part of the decay curve appeared always to be proportional to the initial degree of saturation. This linear part can be represented by

$$\alpha(t) = \alpha(0) - kt \qquad (5.7)$$

in which k is the rate constant of this zero-order process. The rate constants obtained in various experiments are presented in Fig. 5.3 as a function of the salt concentration. Within experimental accuracy the rate constant is

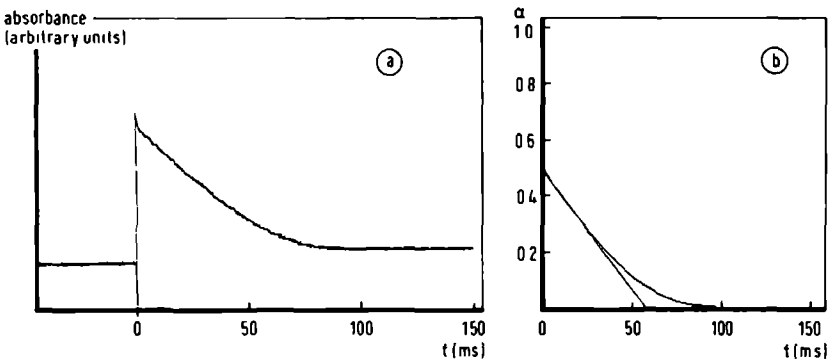



Fig 5 2 Dissociation curve for the complex of poly(dA) and gene-5 protein at an initial degree of saturation of 0 5 Dissociation is effected by increasing the concentration of NaCl to 2 M The total nucleotide concentration is 100 μM A Experimental decay of the absorbance B. Decay of the degree of saturation (α) and the linear curve which fits a large part of the experimental time course

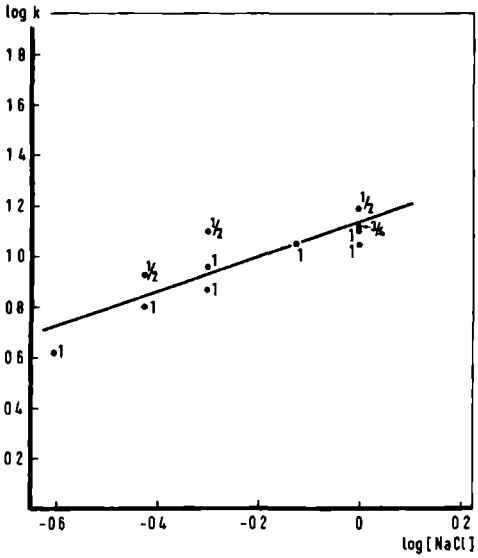



Fig. 5.3. Apparent dissociation rate constant k as a function of the final sodium chloride concentration and the initial degree of saturation for gene-5 protein - poly(dA) complexes. All data points are labeled with the corresponding initial degree of saturation. These points reflect values averaged over about 15 experiments. The final concentration of poly(dA) amounts 50 μM. The data are obtained from absorbance and fluorescence measurements.

independent of the initial degree saturation of the polynucleotide. It follows from Fig. 5.3 that the ionic strength dependence of k, i.e. d log k / d log [NaCl] = 0.68, is rather low compared with the dependence of the intrinsic binding constant, $K_{int}$, on the salt concentration: d log $K_{int}$ / d log [NaCl] = -4.0 (see [Chapter 3 of this thesis]).

The results obtained for the dissociation kinetics of gene-5 protein - M13 DNA complexes are quite different from those obtained for the poly(dA) complexes. M13 DNA exhibits only a linear decay curve when the DNA is initially completely saturated with protein ($\alpha(0) = 1$). However, the rate constant k is

much smaller than for the poly(dA) complex: at 1 M NaCl the relaxation time $\tau$ = 2.2 s, which means that log k = -0.34, while the corresponding value for the poly(dA) complex is +1.07. This makes k about 25 times smaller for M13 DNA at $\alpha(0)$ = 1 compared with poly(dA) at the same conditions. Moreover decay curves are obtained at $\alpha(0)$ values smaller than unity which can be described by a mono-exponential function. Values for the relaxation time $\tau$ thus obtained are 0.56 ± 0.04 s at $\alpha(0)$ = 0.75 and 0.42 ± 0.02 s at $\alpha(0)$ = 0.50, both in 1 M NaCl.

5.4b Effect of binding of gene-5 protein on the $^1$H NMR spectra of polynucleotides

For the detection of a possible influence of gene-5 protein binding on the NMR spectra of polynucleotides, especially concerning the line width of the nucleotide resonances, a solution containing the protein was added to a poly(dA) or a poly(dT) solution of 2 mM and spectra were recorded as a function of the degree of saturation of the polynucleotides with protein. These spectra are collected in Fig. 5.4. The spectra at the bottom of this figure are those of the polynucleotide without protein. The poly(dT) spectra (right) contain the resonances of the base H6 proton (7.50 ppm), the H1' proton (6.08 ppm), the H3' proton (obscured by the residual water peak at 4.7 ppm), the H4', H5'/H5" protons at 3.98 to 4.16 ppm, the H2'/H2" protons at 2.17 to 2.37 ppm, the methylproton at 1.73 ppm and a buffer peak at 1.55 ppm. The poly(dA) spectra (on the left) show the resonances of the H8 proton at 7.77 ppm, the H2 proton at 7.19 ppm, the H1' proton at 5.41 ppm, the H4', H5'/H5" protons around 4 ppm, the H2'/H2" protons at 2.15 to 2.31 ppm and a buffer peak at 1.55 ppm. It is obvious that these resonances are hardly broadened by addition of protein. However, a strong decrease of the signal to noise ratio is observed at higher protein concentration (see Fig. 5.4B-E). At a degree of saturation of the nucleotide with protein of 0.77 (Fig. 5.4E) the nucleotide resonances have disappeared completely from the spectrum and only the buffer peak at 1.55 ppm and some protein resonances are visible.

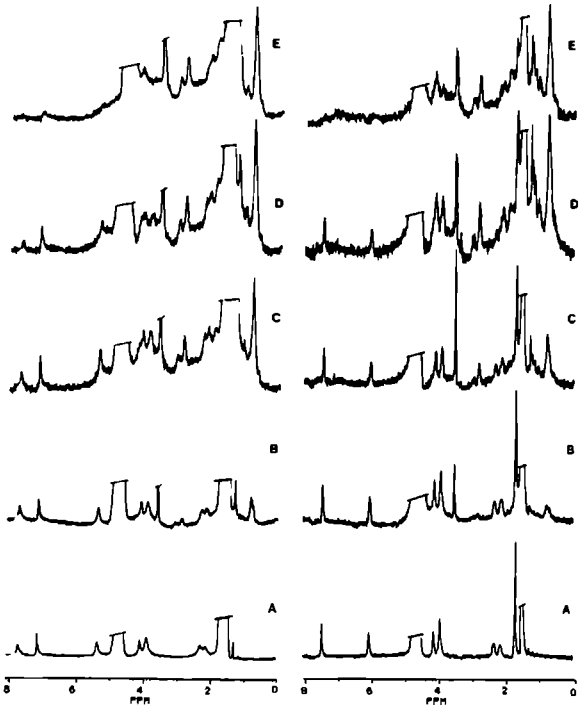



Fig. 5.4. Proton NMR spectra at 300 MHz of poly(dA) (left) and poly(dT) (right) to which gene-5 protein is added For the poly(dA) spectra the degree of saturation is equal to (a) 0, (b) 0 29, (c) 0.49, (d) 0 63 and (e) 0 78, and for the poly(dT) spectra (a) 0, (b) 0 24, (c) 0 50, (d) 0 64 and (e) 0.77.

## 5.5 DISCUSSION

The most remarkable difference between the dissociation kinetics of complexes of gene-5 protein with M13 DNA and poly(dA) is the occurrence of zero-order kinetics for the poly(dA) complex at all initial degrees of saturation studied, while for the M13 DNA this zero-order kinetics is only observed when the

degree of saturation equals unity. At smaller $\alpha(0)$ values the M13 DNA complex displays a single-exponential decay. The latter dissociation behaviour is in fact similar to that reported by Lohman [94] for complexes of T4 gene-32 protein with various homopolynucleotides. The dissociation kinetics of these complexes could be explained by a model in which the dominant mode of dissociation is the release of single contiguously bound protein, i.e dissociation from the ends of protein clusters, while at the same time it is assumed that no redistribution of bound protein takes place during the dissociation process. It was assumed [94] that the absence of redistribution is confirmed by the observation of zero-order kinetics in the dissociation experiments at $\alpha(0) = 1$, but, as we shall show below, this conclusion does not hold universally. In the model developed by Lohman [101] the following equations are used to describe the dissociation time course provided that the cooperativity of binding is high enough ($\omega = 500$ or larger), which is valid for gene-32 protein binding as well as for gene-5 protein binding. At initial complete saturation the concentration of cluster ends is twice as large as the concentration of polynucleotide molecules, so that the rate of the dissociation is given by

$$-(d/dt)\ [P_b] = 2\ k_s\ [CL_t(0)] = 2\ k_s\ [P_b(0)](n/M) \qquad (5.8)$$

which leads to

$$[P_b(t)] = [P_b(0)] - 2\ k_s\ [P_b(0)](n/M)t \qquad (5.9)$$

where $[CL_t(0)]$ is the total concentration of protein clusters at time zero, $k_s$ is the rate constant for the dissociation of single contiguously bound protein from a cluster end, $[P_b]$ is the concentration of bound protein, n is the number of nucleotides occluded by the protein and M is the number of nucleotides which form one polynucleotide chain. Eq. 5.9 shows indeed the linear relationship between the amount of bound protein and the time (cf. Eq. 5.7). At $\alpha(0)$ values smaller than unity the time course for the dissociation of bound protein is given by

$$-(d/dt)\ [P_b] = 2\ k_s\ (1 - p(0))[P_b] \qquad (5.10)$$

where (1 - p(0)) represents the fraction of cluster ends which are present initially [94,101]. The dissociation processes which can be described by Eq. 5.10 give rise to single exponential decay curves. Fig. 5.5 gives the dissociation curves calculated according to the Lohman model for the gene-5 protein - M13 DNA complexes with n = 4 and a cooperativity factor of $\omega$ = 500. It is clear that a linear decay is only observed for $\alpha(0)$ = 1. If this line (a) is taken to represent the experimental straight line for $\alpha(0)$ = 1 with an observed linear decay time of 2.2 s, then the experimental curve observed for $\alpha(0)$ = 0.75 with an exponential decay time of 0.56 s is given by curve (b). Comparison with the calculated curve (c) for $\alpha(0)$ = 0.75 reveals that the actual dissociation of the complex proceeds much slower than the Lohman model predicts. This means that this model is not applicable to the dissociation kinetics of gene-5 protein - M13 DNA complexes at $\alpha(0)$ values below unity.

The dissociation of gene-5 protein - poly(dA) complexes is characterized by linear decay curves also at $\alpha(0)$ smaller than 1. This behaviour cannot be understood from Lohman's model unless the value of the cooperativity factor is so high that the protein is always bound in one cluster per nucleotide chain, even at a low degree of saturation of the polynucleotide. However, for gene-5 protein $\omega$ is about 500 and is therefore much smaller than the cooperativity factor of the gene-32 protein (between 2000 and 10,000), for which the Lohman model has proved to be valid [94]. A simple explanation for the behaviour of gene-5 protein complexes could be the occurrence of a reorganisation of bound protein clusters during the dissociation process. Then Lohman's condition of absence of redistribution is not fulfilled. If we assume thatthis rearrangement is a very fast process, so that at each moment an equilibrium exists for the distribution of the bound protein over the lattice in accordance with the theory of McGhee and Von Hippel [71], then we can use Epstein's theory [93] for the construction of the irreversible dissociation curves. For the gene-5 protein complex (n = 4 and $\omega$ = 500) these are given in Fig 5.6. In Epstein's model the dissociation of isolated bound protein molecules is predominant, especially at a low degree of saturation where they are present in sufficient amount, because their binding constant is about 500 times weaker

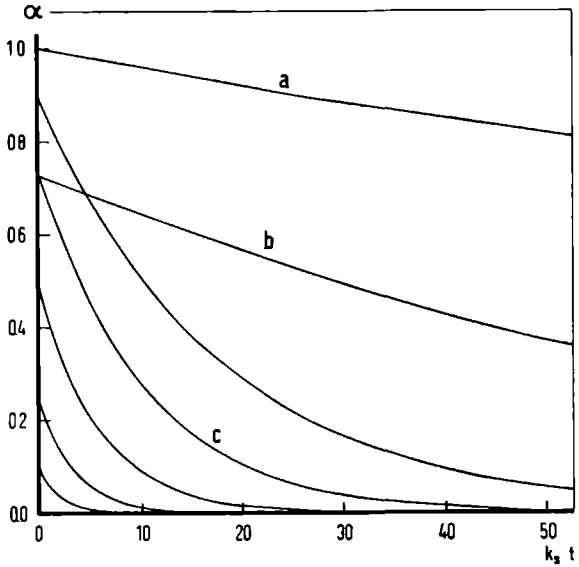



Fig 5 5 Dissociation curves according to the non redistribution dissociation model, given by Lohman [94], see text These curves are calculated using Eqs 5 16 and 5 17, with n = 4 and ω = 500 The curves which are experimentally obtained for the dissociation of gene-5 protein - M13 DNA complexes are also added A linear time constant of 2 2 s was found for the experiment in which α(0) = 1 and the corresponding curve was rescaled in order to make it coincide with the curve (a) calculated according to the Lohman model A mono exponential curve was measured at α(0) = 0 75 with a time constant of 0 56 s This curve (b) was rescaled identical to the rescaling procedure for curve (a) The curve (c) is the mono exponential curve calculated with the Lohman model at α(0) = 0 75

than for single contiguously bound molecules. Therefore the curves in Fig. 5.6 show in the beginning of the dissociation process only a very slow decay, especially at α(0) values close to unity. When the dissociation proceeds its rate drastically increases. So in this case no linear decay curves can be ob-

tained. This makes the Epstein model, in which an infinitely fast redistribution is assumed, not applicable for the description of the dissociation of gene-5 protein - poly(dA) complexes.

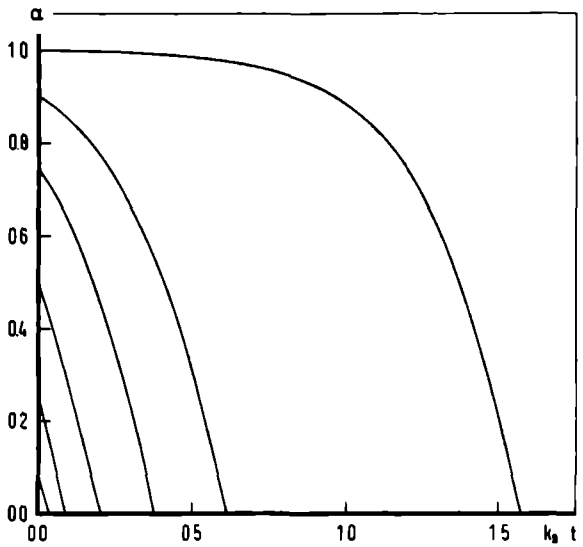



Fig. 5.6. Dissociation curves according to the IELE (Instantenous Establishment of Lattice Equilibrium) dissociation model, given by Epstein [93], see text. Relevant parameters are the same as those taken in Fig. 5.5, i.e. $n = 4$ and $\omega = 500$.

However, as will be shown now, it is still possible to explain the behaviour of the dissociation kinetics in this case from the occurrence of a redistribution of bound protein during the dissociation, if the following assumptions are valid: (1) the dissociation of bound protein is only possible from isolated binding sites. The dissociation of single or double contiguously bound protein is too slow to be detectable. (2) The rate of the association reaction is very low and can be neglected with respect to the dissociation rate. (3) During the dissociation rearrangement of bound protein clusters continuously takes place including the formation of newly isolated bound protein i.e. clusters consisting of only one protein molecule. (4) A steady state situation is reached in which

the concentration of isolated bound protein is constant in time. This steady state situation does not occur at the beginning or at the end of the dissociation time course

For the description of the total process then two successive equilibria are important. First we consider the redistribution of clusters of size q, $CL_q$. In this process a cluster of size one, $CL_1$, is formed (forward reaction with rate constant $k_r$) or bound to a larger cluster (backward reaction with rate constant $k_{r'}$):

$$x\ CL_1 + \Sigma_{q=2}^{\infty} CL_q \underset{k_{r'}}{\overset{k_r}{\rightleftarrows}} (x+1)\ CL_1 + \Sigma_{q=2}^{\infty} CL_q' \qquad (5.11)$$

Irrespective of the value of x (i.e. the number of isolated bound protein molecules per nucleotide chain, which is not necessarily the same for each polynucleotide) the net reaction can be written as

$$D \underset{k_{r'}}{\overset{k_r}{\rightleftarrows}} P_1 + D' \qquad (5.12)$$

in which $P_1$ equals $CL_1$, the isolated bound protein. D is used as a notation for $\Sigma_q\ CL_q$ (q = 2, 3, .. $\infty$). Secondly, dissociation of isolated bound protein molecules as well as association of free protein molecules ($P_f$) with $S_1$ may occur, where $S_1$ is an empty binding site which is not contiguous to a bound protein molecule. This part of the process can be described by

$$P_1 \underset{k_{1'}}{\overset{k_1}{\rightleftarrows}} P_f + S_1 \qquad (5.13)$$

in which $k_i$ and $k_{i'}$ are the dissociation and association rate constants. In the steady-state approximation the time-dependent changes of [D] and $[P_i]$ or $[P_b]$ are equal:

$$-(d/dt)\ [D] = -(d/dt)\ [P_i] = -(d/dt)\ [P_b] \qquad (5.14)$$

Furthermore it follows from Eqs. 5.12 and 5.13 that

$$-(d/dt)\ [D] = k_r\ [D] - k_{r'}\ [D'][P_i] \qquad (5.15)$$

$$-(d/dt)\ [P_i] = k_i\ [P_i] - k_{i'}\ [P_f][S_i] \qquad (5.16)$$

Combination of these three equations yields

$$-(d/dt)\ [P_b] = (k_r k_i[D]) / (k_i + k_{r'}[D']) \qquad (5.17)$$

According to criterion number 2 (see above) the association rate constant $k_{r'}$ is assumed to be zero. Subsequently we consider the dissociation reaction and assume that the rearrangement is fast compared with the dissociation, i.e. $k_{r'}[D'] \gg k_i$. Hence Eq. 5.17 can be converted using $k_r / k_{r'} = [D'][P_i] / [D]$ (see Eq. 5.12) into

$$-(d/dt)\ [P_b] = (k_r k_i[D]) / (k_{r'}[D']) \qquad (5.18)$$

or

$$(d/dt)\ \alpha = (d/dt)\ n[P_b] / [N_t] = -k_i n[P_i] / [N_t] \qquad (5.19)$$

From Eq. 5.19 the features which characterize the dissociation time course of complexes of gene-5 protein with poly(dA), as described in the Results section, can be easily explained.
Since the system finds itself in a steady state situation $[P_i]$ is constant which means that the reaction rate is also constant and the reaction is zero order, as found for the dissociation reaction of gene-5 protein complexes with poly(dA). The breakdown of the steady state approximation in the last stage of the dissociation, resulting in a deviation from linearity, has also been observed (see Figs. 5.1 and 5.2). The invalidity of the steady state approximation in the first part of the dissociation process of these complexes will result in a non-linear decay of the curves in the beginning of the dissociation. However, this effect cannot be observed experimentally due to the dead time of the stopped flow apparatus.
The steady state condition that $[P_i]$ does not alter during the dissociation process, i.e. upon lowering the degree of saturation of the nucleotide lattice with protein, implies that the dissociation rate is also independent on the initial degree of saturation, see Eq. 5.19. This has indeed been found for gene-5 protein - poly(dA) complexes (see Fig. 5.3). Therefore this result is an indication of the validity of the assumptions made above.

It follows from the discussion conducted sofar that neither with Lohman's nor with Epstein's model the dissociation behaviour of the gene-5 protein - M13 DNA complexes can be described. The dissociation kinetics of these complexes could only be explained from Lohman's non-redistribution model at $\alpha(0) = 1$. At lower $\alpha(0)$ values this model failed to describe the time course properly. However, the redistribution model outlined above may also account for the dissociation kinetics of these complexes at $\alpha(0)$ values smaller than or equal to unity. To this end we have to take into consideration the observation that the value of the equilibrium constant for the binding of a gene-5 protein to M13 DNA depends on the base composition of the sequence to which it is bound (see Chapter 3 of this thesis). Because the binding of gene-5 protein to poly(dT) is much stronger than for example to poly(dA) it can be expected that binding to sequences rich in thymidines is very tight compared with the

binding to stretches rich in adenines. As a consequence the redistribution of bound ligands will now be much slower than for the gene-5 protein - poly(dA) complexes. Hence the assumption that the rearrangement is fast compared with the dissociation reaction ($k_r \cdot [D'] >> k_l$), which is used to derive Eq. 5.19, may no longer hold. On the contrary, if $k_r \cdot [D'] << k_l$, which means that the rearrangement is much slower than the dissociation itself, then Eq. 5.17 transfers to the following expression:

$$-(d/dt)\ [D] = -(d/dt)\ [P_b] = k_r\ [D] \qquad (5.20)$$

At $\alpha(0) = 1$ (no empty binding sites) the rearrangement is very slow and the dissociation follows the kinetics described by Lohman. At $\alpha(0) < 1$ Eq. 5.20 predicts the exponential decay which was found experimentally for the dissociation of gene-5 protein - M13 DNA complexes. The relaxation times found in this situation ($\tau = 0.56 \pm 0.04$ s at $\alpha(0) = 0.75$ and $\tau = 0.42 \pm 0.02$ s at $\alpha(0) = 0.50$, both in 1 M NaCl) have to be interpreted according to Eq. 5.20 as the redistribution relaxation time $k_r^{-1}$. This indicates that the rate of the rearrangement indeed increases upon lowering the initial degree of saturation.

In summary it appears that a dissociation model in which the protein is assumed to dissociate from isolated sites on the polynucleotide chain, accompanied by a redistribution of the bound protein, can account for the dissociation kinetics of the complexes of gene-5 protein with poly(dA) as well as M13 DNA without regard to the initial degree of saturation. This model does not include the restriction of the existence of a lattice equilibrium according to the theory of McGhee and Von Hippel [71] at any moment during the dissociation, as has been introduced in the model of Epstein. The redistribution indicates that a translocation takes place along the chain, like sliding, i.e. a one dimensional diffusion along the lattice and (or) hopping, i.e. microscopic dissociation and association to correlated areas of the backbone. It can be noted that the kinetic experiments with gene-5 protein and fd DNA, performed by Shimamoto and Utiyama [59] have provided evidence for the correctness of the assumption that dissociation of bound protein most likely occurs from isolated sites. The difference between the dissociation kinetics of

gene-32 protein complexes (Lohman model) and gene-5 protein complexes might be caused by the difference in the magnitude of the cooperativity factor, i.e. $\omega \approx 500$ for gene-5 protein and $\omega \approx 5000$ for gene-32 protein [94]. In that case a lower cooperativity should facilitate the redistribution.

The salt concentration dependency of the dissociation rate constant, $d \log k_d / d \log [NaCl]$, equals 0.68 for the dissociation of gene-5 protein - poly(dA) complexes. This is much less than the corresponding salt concentration dependency of the intrinsic binding constant: $d \log K_{int} / d \log [NaCl] = -4.0$ (see Chapter 3 of this thesis). It indicates that the salt concentration mainly influences the association rate: $d \log k_a / d \log [NaCl] = -3.32$. This effect is quite similar to that found, for example, for the double helix formation of DNA and RNA [20, 21]. Here it appears that also in particular the helix recombination rate is influenced by the ionic strength.

The NMR experiments show that addition of gene-5 protein to poly(dA) or poly(dT) hardly affects the linewidth of the nucleotide resonances, but on the other hand causes a drastic decrease of the signal to noise ratio of these resonances. This is in sharp contrast to the results of Römer et al. [145] who observed a strong broadening of the nucleotide signals upon addition of Escherichia coli single stranded DNA binding protein (SSB) to poly(dA) or poly(dT), while the signal to noise ratio remained almost unchanged. From this broadening they concluded to a fast exchange of SSB between different binding sites, indicating a high rate of translocation of SSB on the polynucleotide. For the gene-5 protein - poly(dA) complexes the stopped flow experiments also point to a fast translocation of the protein from protein clusters to isolated sites on the lattice. The reason why this does not result in line broadening of the nucleotide resonances in this case must be found in the high cooperativity of the gene-5 protein binding. As a consequence of this the concentration of protein bound on isolated sites is very low and the major part of the protein is bound in a cluster. Since the isolated bound protein molecules participate in the rearrangement and therefore could give rise to broadening of the nucleotide resonances, this line broadening will not be observed in this case. To illustrate this we have calculated the fraction of gene-5 protein present as isolated, single contiguously or double contiguously bound protein as a function of the degree of saturation, using the theory of

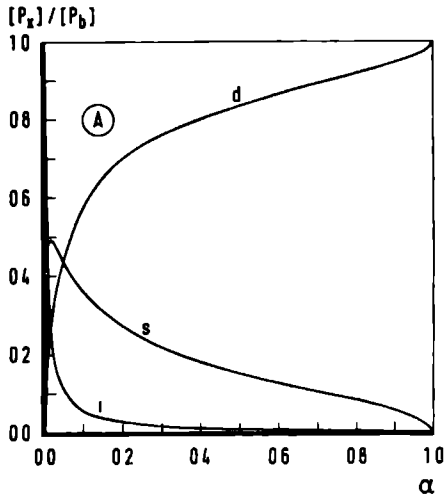



Fig 5 7 Distribution of protein on a polynucleotide as derived from the theory of McGhee and Von Hippel [71] for cooperative binding including binding site overlap The fraction of the total concentration of bound protein ($[P_b]$),which is bound isolated single contiguously or double contiguously (x = 1, s or d respectively) is given as a function of the degree of saturation of the lattice For the calculations the values n = 4 and ω = 500 are used

McGhee and Von Hippel [71] and the binding parameters as determined from fluorescence titrations [Chapter 3 of this thesis]. These results are given in Fig. 5.7, which clearly shows the small fraction of isolated bound protein at high and medium degrees of saturation. Finally, the observed decrease of the signal to noise ratio at higher degrees of binding must be a consequence of the immobilization of nucleotide stretches after binding a (large) protein cluster.

# CHAPTER 6

# FURTHER CHARACTERIZATION OF THE DIFFERENT BINDING MODES FOR POLY- AND OLIGODEOXYNUCLEIC ACIDS

## 6.1 ABSTRACT

The binding of gene-5 protein, encoded by bacteriophage M13, to oligodeoxynucleic acids was studied by means of fluorescence binding experiments, fluorescence depolarisation measurements and irreversible dissociation kinetics of the protein - nucleotide complexes with salt. The binding properties thus obtained are compared with those of the binding to polynucleotides, especially at very low salt concentration. It appears that the binding to oligonucleotides is always characterized by a stoichiometry n of 2 to 3 nucleotides per protein, and the absence of cooperativity. In contrast the protein can bind to polynucleotides in two different modes, one with a stoichiometry n = 3 in the absence of salt and another with n = 4 at moderate salt concentrations. Both modes have a high intramode cooperativity ($\omega$ about 500) but are non-interacting and mutually exclusive For deoxynucleic acids with a chainlength of 25 - 30 residues a transition from oligonucleotide to polynucleotide binding is observed at increasing nucleotide to protein ratio in the solution. The n = 3 polynucleotide binding is very sensitive to the ionic strength and only detectable at very low salt concentrations. The ionic strength dependency per nucleotide of the n = 4 binding is much less and comparable with the salt dependency of the oligonucleotide binding. Furthermore it appears that the influence of the salt concentration on the oligonucleotide binding constant is to about the same degree determined by the effect of salt on the association and dissociation rate constants Model calculations indicate that the fluorescence depolarisation titration curves can only be explained by a model for oligonucleotide binding in which a protein dimer binds with its two dimer halves to the same strand.

In addition it is only possible to explain the observed effect of the chain length of the oligonucleotide on both the apparent binding constant and the dissociation rate by assuming the existence of interactions between protein dimers bound to different strands. This results in the formation of a complex consisting of two nucleotide strands with protein in between and stabilized by the dimer - dimer interactions.

## 6.2 INTRODUCTION

It is now well-known, at least for prokaryotes, that single stranded DNA binding proteins participate in the process of DNA replication. The gene-5 protein encoded by the filamentous bacteriophage M13 is a case in point. It binds strongly and cooperatively to single stranded DNA and is able to destabilize double stranded DNA. During DNA metabolism it is instrumental in the induction of a switch from replicative form replication to single strand DNA synthesis.

Determinant in the DNA protein interaction is the stoichiometry and the nature of the binding The question of the stoichiometry of the gene-5 protein DNA binding, i.e. the number of nucleotides covered by one protein molecule has been addressed in a number of studies [33,35,37,45,59,60,Chapter 3 of this thesis] and has been reviewed recently by Kansey et al [70]. A so-called n = 4 binding mode, in which the protein molecule covers four nucleotides, is considered to be the normal binding state of the gene-5 protein. Interestingly, at some conditions the protein can be bound in a so-called n = 3 binding mode, which is clearly distinct from the more common binding state mentioned above. The n = 3 binding mode was first discovered in a series of NMR experiments in which the complex formation between the gene-5 protein and a series of oligo-(dA)s was studied [67] and henceforth is called the oligonucleotide binding mode, in contrast to the common n = 4 binding mode which was found for gene-5 protein polynucleotide complex formation and therefore is called the polynucleotide binding mode. Subsequently, Gray and collaborators showed however, that at low salt conditions the gene-5 protein is also able to bind to polynucleotides in the n = 3 binding mode [70].

The data available sofar suggest that n = 3 binding to oligonucleotides exhibits binding features different from n = 3 binding to polynucleotides. For instance complex formation of the gene-5 protein with oligonucleotides seems to be almost non-cooperative [67] while model calculations suggest that binding to polynucleotides is highly cooperative in both binding modes [70].
In an attempt to obtain a unified picture for the complex formation of the gene-5 protein with DNA we further explored the binding characteristics of the protein, especially those concerning the binding to oligonucleotides. In the present chapter results from fluorescent binding and fluorescent depolarisation studies are discussed. Moreover the kinetics of the irreversible dissociation of gene-5 protein - oligo(dT) complexes with salt, as follows from stopped-flow experiments, is used in the interpretation of these results and is compared with the corresponding kinetics for gene-5 protein - polynucleotide complexes, which is described in the preceding chapter.

## 6.3 MATERIALS AND METHODS

### 6 3a Materials

Gene-5 protein was isolated and purified as described previously [60], except that during sonication of E. coli no Dnase I was added. Concentrations of the gene-5 protein were measured using a molar absorption coefficient at the absorption maximum (276 nm) of 7100 $M^{-1}$ $cm^{-1}$ [29].
The oligo(deoxythymidylic acid)s and oligo(deoxyadenylic acid)s with 5'-terminal phosphates were purchased from P-L Biochemicals. Concentrations were measured using the following absorption coefficients: $\varepsilon$ = 8440 $M^{-1}$ $cm^{-1}$ (266 nm) for $(dT)_6$, $\varepsilon$ = 9980 $M^{-1}$ $cm^{-1}$ (257 nm) for $(dA)_8$, $\varepsilon$ = 8690 $M^{-1}$ $cm^{-1}$ (266 nm) for $(dT)_8$, $\varepsilon$ = 8600 $M^{-1}$ $cm^{-1}$ (266 nm) for $(dT)_{12}$, $\varepsilon$ = 9910 $M^{-1}$ $cm^{-1}$ (257 nm) for $(dA)_{12}$, $\varepsilon$ = 8540 $M^{-1}$ $cm^{-1}$ (266 nm) for $(dT)_{18}$, $\varepsilon$ = 9910 $M^{-1}$ $cm^{-1}$ (257 nm) for $(dA)_{16}$ and $\varepsilon$ = 8570 $M^{-1}$ $cm^{-1}$ (265 nm) for $(dT)_{25\text{-}30}$. The first three absorption coefficients were obtained directly from the literature [137] and the others were found after interpolation from the data in this reference. $(dT)_{25\text{-}30}$ is a mixture of oligo(deoxythymidylic acid)s ranging in length from 25 to 30 residues.

For poly(dA) $\varepsilon$ = 9300 $M^{-1}$ $cm^{-1}$ (260 nm) and for poly(dT) $\varepsilon$ = 8700 $M^{-1}$ $cm^{-1}$ (265 nm).
M13 DNA was purified according to Konings and collaborators; a molar absorption coefficient of 7400 $M^{-1}$ $cm^{-1}$ at 259 nm was used for this compound [106].

Throughout this chapter the concentration of nucleotides is expressed as the number of moles mononucleotide per liter. All nucleotides were used as the $Na^+$ salts. The measurements were performed in a buffer containing 10 mM sodium cacodylate and 50 mM sodium chloride (unless otherwise mentioned) at pH 6.9.

### 6.3b Fluorescence binding experiments

Fluorescence binding measurements were performed with a Perkin Elmer fluorescence spectrophotometer MPF-4 equipped with a 150 W Xenon lamp. The instrument was used in the so-called ratio mode; this means that relative fluorescence intensities were measured instead of absolute quantum yields. Excitation was at the absorption maximum (276 nm) of the tyrosyl residues and the fluorescence was measured at the emission maximum (303 nm). The fluorescence intensities were corrected for the inner filter effect, contribution of the water Raman signal and for dilution during titration. Fluorescence titration experiments were performed according to the method described previously [37,Chapter 3 of this thesis]. All measurements were carried out at 5 $^oC$.

### 6.3c Fluorescence depolarisation measurements

Time-resolved fluorescence experiments were performed at 20 $^oC$ on the EMBL fluorescence spectrometer, which is described in detail elsewhere [138]. The system consists of a synchronously pumped cavity dumped dye laser, the output of which is frequency doubled to 285-310 nm. The pulse width of the system is 10 ps, with a repetition rate of 4 MHz. The tyrosyl fluorescence of the sample was excited at 285 nm and observed through a 20 nm wide inter-

ference filter centered around 300 nm. The time dependent change of the fluorescence anisotropy, upon excitation with a very short plane polarised pulse, was used to determine the rotational correlation time of the protein (see chapter 4).
The time dependent protein fluorescence anisotropy was measured in several titration experiments. Each individual point in a titration experiment was measured with a freshly prepared sample to avoid artifacts due to protein irradiation damage. During the titration either the protein concentration or the nucleotide concentration was kept constant. The data were analysed using a non-linear least squares fitting procedure, [142].

### 6.3d Stopped flow irreversible dissociation experiments

To gene-5 protein - oligo(dT) complexes at low ionic strength a concentrated salt solution was added to dissociate the complex. Rapid mixing was obtained in a modified Durrum stopped-flow equipment and the reaction was monitored by the disappearance of the fluorescence quenching upon dissociation. The method is described in detail in the preceding chapter. The experiments were conducted at 8 $^{o}$C, unless otherwise mentioned.

## 6.4 RESULTS

### 6.4a Binding parameters from fluorescence titrations for the binding to polynucleotides

The binding of gene-5 protein to different polynucleotides at different salt concentrations (down to 50 mM NaCl) is characterized by an occupation of four nucleotides by one protein monomer at complete saturation. At the low salt concentration, i.e. at about 50 mM NaCl, a high value of the binding constant is measured. This implies that in a titration during the first part of the experiment virtually all added protein is bound to the polynucleotide. This results in a linear increase of the fluorescence until complete saturation of the polynucleotide is obtained. Upon further addition of protein again a linear increase of the fluorescence is observed, but with a larger increment because now the fluorescence of the protein is not quenched by binding.

From the intersection of the two straight lines (thus obtained) the stoichiometry of the binding (n) can be easily derived, while the ratio of the slopes of the two lines gives the fraction by which the fluorescence of the protein is quenched upon binding ($Q_{max}$) [37]. In order to derive values for the binding constant, titrations have to be performed at somewhat higher salt concentrations, where the effective binding constant, $K_{int}\omega$, is lower and the binding isotherms have a sigmoidal shape. ($K_{int}$ characterizes the complex formation of one protein to an isolated binding site on the lattice, and the cooperativity factor, $\omega$, defines the factor by which the binding constant increases when a protein binds adjacent to an already bound protein relative to isolated binding). The latter feature demonstrates the presence of the high cooperativity of the binding, which is not manifest at low salt conditions (50 mM NaCl). A rather different behaviour is observed at significantly lower salt concentrations than 50 mM NaCl. Fig. 6 1 shows the kind of fluorescence titration curves measured at very low ionic strength (the buffer contained only 1 mM sodium cacodylate). Titration curves obtained for M13 DNA and poly(dA) (curve a and b in Fig. 6.1) exhibit two intersection points; the same results were obtained for poly(dT) (not shown). This indicates the presence of two different binding modes in agreement with results obtained from circular dichroism experiments by Kansey et al. [70]. It is noted in passing that at the salt concentrations considered the binding is very strong so that the apparent shallow form of the titration curves cannot be attributed to weak non-cooperative binding. From a linear extrapolation of the fluorescence at low protein over nucleotide (P/N) ratios and the fluorescence at high P/N ratios a stoichiometry of n = 4 is derived (dashed curves, Fig. 6.1a and b), which indicates that this binding mode dominates at low P/N values. The other mode, which has a stoichiometry of n = 3 as follows from the second intersection point at P/N = 1/3, is found instead at higher protein over nucleotide ratios. Both modes have different $Q_{max}$ values as well, i.e. $Q_{max}$ = 0.60 for the n = 4 binding mode and $Q_{max}$ = 0.45 for the n = 3 binding mode. This lower quenching factor in the n = 3 mode leads to a larger increase of the fluorescence intensity in the titration curve between P/N = 0.2 and 0.4. As a result both modes are detectable in the same curve. This is clearly demonstrated by the simulated curve c (solid curve) in Fig. 6.1, which was derived with the aid of the afore-mentioned experimental n and $Q_{max}$ values. At a stoichiometric binding of just n = 4 the lower dashed line would have been measured while if only n = 3 binding were possible the isotherm would follow

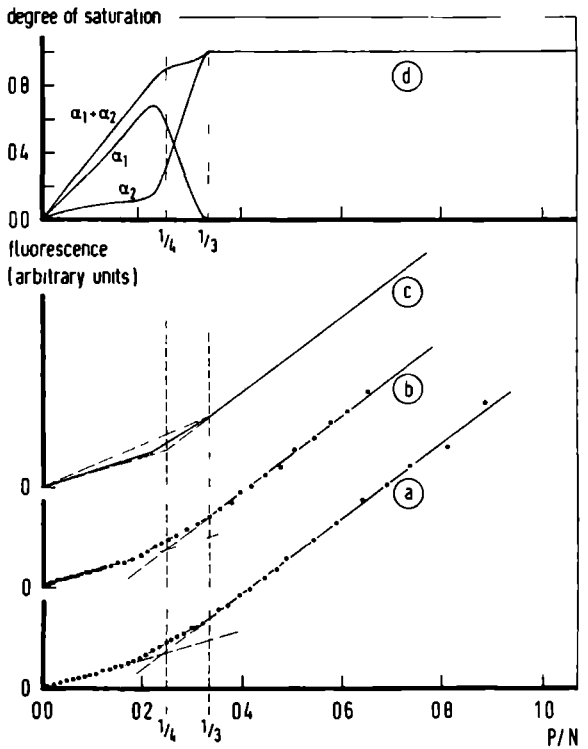



Figure 6 1 Fluorescence titration curves for the titration of M13 DNA (curve a) and poly(dA) (curve b) with gene-5 protein Curve c is a calculated curve for the situation of dual mode binding with stoichiometries of four (n = 4) and three (n = 3) nucleotides per protein monomer For the calculation binding parameters were used as derived from the curves a and b In the insert d the degree of saturation of the polynucleotide with protein in the n = 4 binding mode ($\alpha_1$) and in the n = 3 binding mode ($\alpha_2$) are presented as a function of the P/N raratio These simulations were performed with the same binding parameters as used for the calculation of curve c The dashed vertical lines indicate positions where P/N equals the reciprocal of the stoichiometric values n = 4 and n = 3 respectively The dashed parts of curve c indicate the change of the fluorescence expected when only n = 3 or n = 4 binding would occur The titrations were performed at pH 6 9 and 5 °C for solutions containing 27 3 µM M13 DNA (a) or 33 5 µM poly(dA) (b) and also 1 mM sodium cacodylate

the upper dashed line. The degree of saturation corresponding with the fluorescence quenching exhibited in curve c is presented in Fig. 6.1d (see Discussion, section 6.5d). The interpretation presented so far is only valid when for P/N exceeding 1/3 (where the polynucleotide is completely saturated with protein) the same fluorescence intensity is found irrespective of the mode in which the protein is bound. This situation obtains when the ratio of $Q_{max}$ (n = 3) and $Q_{max}$ (n = 4) equals 3/4, which indeed represents the actual situation. This can be checked by considering the expression describing the fluorescence intensity during a titration experiment:

$$F = F_f \{ [P_f] + [P_{b,1}] (1 - Q_{max,1}) + [P_{b,2}] (1 - Q_{max,2}) \} \qquad (6.1)$$

in which the subscripts 1 and 2 denote binding with n = 4 and n = 3 respectively. $F_f$ is the fluorescence per mole of free protein and $[P_f]$ and $[P_b]$ are the concentrations of free and bound protein. In particular the validity of the extrapolation procedure carried out to resolve the n = 4 binding will then be apparent.

### 6.4b Binding parameters from fluorescence titrations for the binding to oligonucleotides

Parameters characteristic for the binding of gene-5 protein to oligodeoxynucleotides were determined in two different types of fluorescence titrations. First, titrations were performed in exactly the same way as already described for polynucleotides, which means that protein was added to an oligonucleotide solution at low salt concentration where the binding is strong. These titrations were carried out for oligo(deoxythymidilic acid)s and oligo(deoxyadenylic acid)s of varying length. In Fig. 6.2 a typical fluorescence titration curve is presented, obtained upon addition of the protein to a $(dT)_{18}$ solution containing 10 mM sodium cacodylate (pH 6 9) and 50 mM NaCl. The number of nucleotides covered by one protein molecule amounts to n = 2.8, while the quenching factor $Q_{max}$ is equal to 0.42. These values are about the same as those obtained for the n = 3 binding mode of polynucleotides. The value of n and $Q_{max}$ determined for a number of oligonucleotides

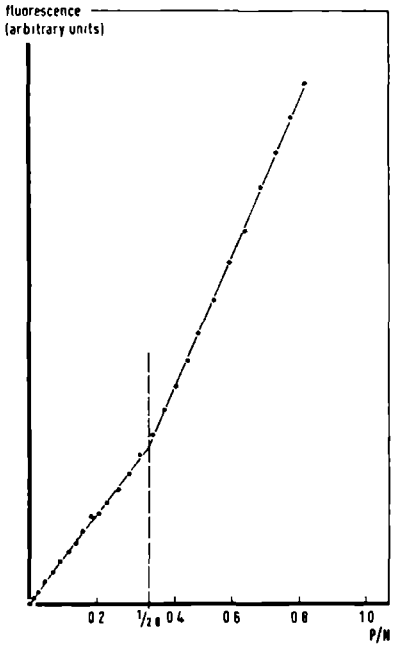



Figure 6 2 Fluorescence titration curve of $(dT)_{18}$ with gene-5 protein at 50 mM NaCl and 10 mM sodium cacodylate, pH 6 9, T = 5 °C The nucleotide concentration varied from 23 19 µM at the beginning to 15 7 µM at the end of the titration

are collected in Table 6.1. It is clear that these results differ systematically from those obtained for the complex formation of gene-5 protein with polynucleotides at the same ionic strength (n = 4 and $Q_{max}$ = 0.60 [29,37,chapter 3 of this thesis]

In the second type of titrations a solution with a high concentration of NaCl is added to a solution which contains a gene-5 protein - oligonucleotide complex with a slight amount of the oligonucleotide in excess, ensuring that at the beginning of the titration, i.e. at low salt concentration, all the protein is bound. During the titration the complex is increasingly dissociated resulting in an enhanced fluorescence intensity. From the observed intensity, $F_{obs}$, the fraction of bound protein, θ, can be calculated with

Table 6.I.
Binding parameters for oligodeoxynucleotide complexes as derived from fluorescence binding isotherms at 5 °C and 50 mM NaCl. Typical nucleotide concentrations during the experiment were in the order of 20 μM.

| oligonucleotide | n | $Q_{max}$ |
|---|---|---|
| $(dT)_6$ | -- *) | -- *) |
| $(dT)_8$ | 2.8 | 0.25 |
| $(dA)_8$ | 2.2 | 0.17 |
| $(dT)_{12}$ | 2.4 | 0.40 |
| $(dA)_{12}$ | 2.3 | 0.41 |
| $(dA)_{16}$ | 2.7 | 0.42 |
| $(dT)_{18}$ | 2.8 | 0.42 |
| $(dT)_{25-30}$ | 2.7 | 0.45 |

*) no binding could be observed

$$\theta = ( [P_t] F_f - F_{obs} ) / [P_t] F_f Q_{max} \qquad (6.2)$$

in which $[P_t]$ is the total protein concentration and $F_f$ is the fluorescence per mole of free protein. Then the fluorescence intensities at the beginning and at the end of the titration are given by $[P_t]F_f(1 - Q_{max})$ and $[P_t]F_f$ respectively. Fig. 6.3 shows salt titration curves recorded for complexes of gene-5 protein with oligo(deoxythymidylic acid)s of a chain length of 8 nucleotides or more, including the complex with poly(dT). It is noted in passing that for the complexes with oligo(deoxyadenylic acid)s (*cf.* Table 6.I) the binding is very weak. This means that they dissociate readily at low salt concentrations, which makes them less suited for this type of titration.

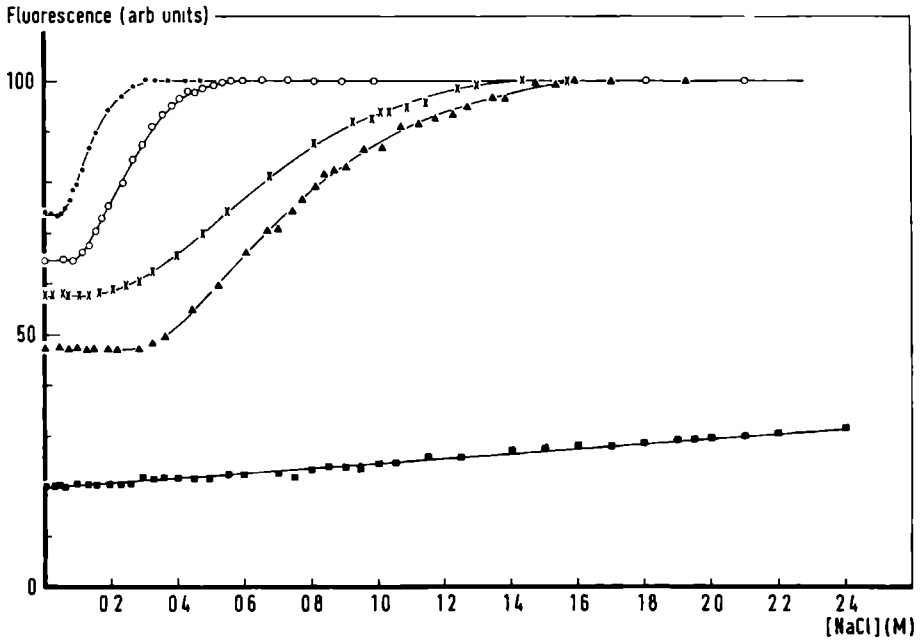



Figure 6 3 Gene-5 protein fluorescence as a function of salt concentration, recorded at 5 °C, upon disruption from $(dT)_8$ (•), $(dT)_{12}$ (o), $(dT)_{18}$ (×), $(dT)_{25-30}$ (▲) or poly(dT) (■) The total protein concentration is about 6 µM

It can be gleaned from Fig. 6.3 that elongation of the chain lengths requires higher salt concentrations to achieve complete dissociation of the complex. Apparently the binding constant increases drastically with the chain length at the same ionic strength. For the polynucleotide, poly(dT), only a small degree of dissociation could be observed, even at the highest concentration of sodium chloride. To analyse the salt titration curves we introduce an apparent binding constant, $K_{app}$, for the binding of gene-5 protein to oligonucleotides, which is dependent on the salt concentration. $K_{app}$ is defined as the reciprocal value of the free protein concentration, $[P_f]$, at half saturation of the nucleotide, i.e

$$K_{app} = [P_{f,1/2}]^{-1} \qquad (6.3)$$

Hence, by adjusting the salt concentration to obtain the situation of half saturation, a binding constant can be deduced. Apparent binding constants, derived in this way from titration curves as given in Fig 6 3, were plotted as a function of the NaCl concentration (Fig. 6.4) The different stability of the various oligonucleotide complexes is reflected by the different values of log $K_{app}$ at the same NaCl concentration in Fig. 6.4. For example, after extrapolation of the straight lines we obtained values for log $K_{app}$ at 1 0 M NaCl of 3.04 for $(dT)_8$, 4 02 for $(dT)_{12}$, 4.78 for $(dT)_{18}$ and 5 09 for $(dT)_{25-30}$. The explanation for the chain length dependency of $K_{app}$ will be given later (see Discussion, section 6.5d).

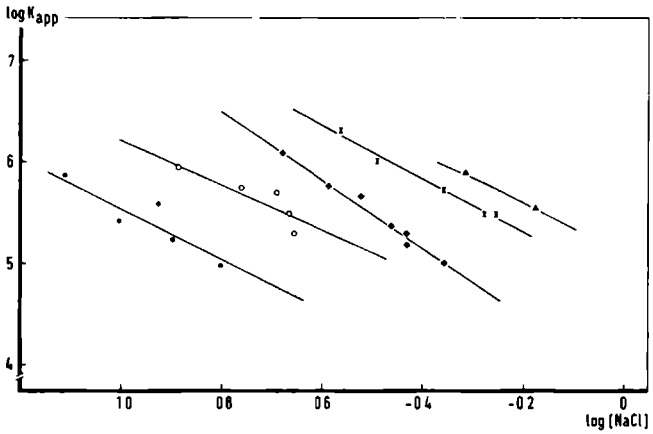



Figure 6 4 Salt dependency of the apparant binding constant $K_{app}$ for gene-5 protein / (deoxythymidilic acid) complexes, as derived from salt titrations of the complex at 5 °C The oligo(deoxythymidilic acid)s consist of 8 (•), 12 (o), 18 (×) or 25-30 residues (▲) For reasons of comparison the curve for the gene-5 protein / M13 DNA complex ($K_{app} = K_{int}\omega$, ◆ ) is also shown

### 6.4c Fluorescence depolarisation experiments

Measurements of the depolarisation of the tyrosyl fluorescence of gene-5 protein at different nucleotide over protein ratios were carried out to determine the rotational correlation time of the protein as a function of this ratio. We have shown previously (see chapter 4) that in this manner it is possible to discriminate between binding with a high or a very low cooperativity factor $\omega$. Thus, after comparison of n, $K_{app}$ and $Q_{max}$ obtained for the different DNA binding modes of gene-5 protein, the fluorescence depolarisation method enables us to investigate the cooperativity connected with these modes. The time dependence of the fluorescence anisotropy of a single particle can be described by the following expression [141].

$$A(t) = A(0)\ e^{-t/\phi} \qquad (6.4)$$

where A(0) is the zero time anisotropy and $\phi$ is the rotational correlation time of the particle. For globular particles we have

$$\phi = f_{rot}\ V_h\ \eta\ /\ kT \qquad (6.5)$$

where $V_h$ is the hydrated particle volume, $\eta$ is the viscosity of the sample and $f_{rot}$ is the relative rotational friction coefficient, which is a measure for the deviation from sphericity. Assuming that the gene-5 protein molecule is spherical we calculate that $\phi$ = 8 00 ns at 20 $^{o}$C for the dimer (M = 19378 g/mole). Measured values of the rotational correlation time are 12.9 ns at a protein concentration of 79 $\mu$M and 15.4 ns at 840 $\mu$M. From these results we concluded that the deviations from the theoretically predicted value are caused by protein aggregation (see chapter 4). Measurements of the time dependence of the fluorescence anisotropy in the presence of oligo- or polynucleotides showed that for all of the complexes studied Eq. 6.4 is applicable, which means that the A(t) versus t curves can be described by a single exponential from which the value of the effective rotational correlation time can

be derived. During titration of the gene-5 protein with oligo- or polynucleotide, complexes with higher molecular weights are formed. This is reflected in a change of the rotational correlation time with the N/P ratio.

Fig. 6.5 shows the effect of complex formation of gene-5 protein with M13 DNA on the protein rotational correlation time at different concentrations of NaCl. The formation of particles with a higher molecular weight is apparent from the increase of $\phi$ with the N/P ratio which proceeds until all the protein is bound at an N/P value of 4. The fluorescence lifetime of tyrosyl residues is relatively short (see chapter 4), which implies that only a narrow time window is available for measuring the fluorescence depolarisation. This means that for the experiments described in this chapter, the higher values of $\phi$ are less accurate and therefore for $\phi$ exceeding 90 ns are represented in the figures by dashed lines, indicating the lower limit of the effective rotational correlation time. In case of M13 DNA the large value of the rotational correlation time obtained for N/P greater than 4 indicates that the protein remains bound in large clusters, even in the presence of an excess of DNA, as prescribed by the McGhee and von Hippel model [71]. The high binding cooperativity precludes a uniform distribution of the bound protein molecules over the available DNA strands (see chapter 4). The convex downward shape of the curve observed in the beginning of the titration can be explained by a model in which a (slow) exchange of protein between the various complexes is assumed (see chapter 4). Similar curves are measured for 93 mM NaCl and 139 mM NaCl solutions. The small differences in the rotational correlation time in the first part of the titration are possibly caused by a difference in the amount of aggregation of the protein samples. Both curves have the same shape and point to a binding stoichiometry of 4 nucleotides per protein monomer at this ionic strength. On the other hand the curve obtained at very low salt concentration, i.e. 1 mM or less, shows a different behaviour. In the first part of the curve a larger increase of the rotational correlation time is observed upon increasing the N/P value. Since we know from fluorescence binding experiments that at all three NaCl concentrations the binding of the protein to DNA is very strong, this different shape cannot be caused by a difference in binding strength at 1 mM NaCl and therefore points to a binding stoichiometry of less than 4 nucleotides per protein monomer at this ionic strength. This is in agreement with the results of the fluorescence titration

experiments which reveal an n = 3 binding mode at very low NaCl concentration. That the stoichiometric value n = 3 is not observed in the depolarisation studies originates from the fact that if both modes are possible, the n = 3 mode will be only observable when the protein is present in excess amount, i.e. at low N/P ratios. Thus, compared with Fig. 6.1, where upon increasing P/N we first observe n = 4 binding followed by n = 3 binding, we now encounter the reverse situation.

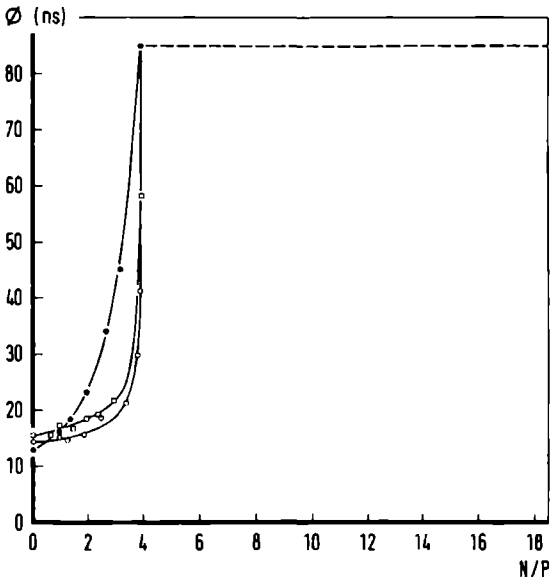



Figure 6.5 Rotational correlation times, $\phi$, of gene-5 protein / M13 DNA mixtures as a function of N/P, the ratio mononucleotide/protein monomer, obtained at 20 °C. The sodium chloride and nucleotide concentrations were kept constant during the titration at less than 1 mM NaCl and 70.47 µM M13 DNA (•), 93 mM NaCl and 177.57 µM M13 DNA (□) or 139 mM NaCl and 220.81 µM M13 DNA (○). The nucleotide concentrations are expressed in mononucleotides. The broken line at higher N/P ratios indicates high values of $\phi$. The exact values cannot be determined because of the fast-decaying protein fluorescence.

Rotational correlation times measured as a function of the N/P ratio for mixtures of $(dT)_{18}$ and gene-5 protein at salt concentrations less than 1 mM are given in Fig. 6.6. The curve is adapted from chapter 4 and is presented here for reasons of comparison. At N/P ≈ 2.5, $\phi$ is at a maximum and upon further increasing of N/P decreases, because the low cooperativity of the binding allows a redistribution of bound protein over the nucleotide lattices when N/P exceeds the stoichiometric value. Hence, in a situation of very strong binding the stoichiometry of the binding follows directly from the N/P value at which the maximum is found. This type of $\phi$ versus N/P curve is only found for low cooperativities (see chapter 4), which can easily be distinguished from the curves obtained for high values of $\omega$ (compare Fig. 6.5).

$\phi$ versus N/P curves measured for the complexation of gene-5 protein with $(dT)_{25-30}$ are also given in Fig. 6.6. Again convex downward curves are obtained in the beginning of the titration. At very low ionic strength the results seem to indicate a stoichiometry of 2 5 to 3 nucleotides per protein monomer. On the other hand at the larger salt concentrations the curve is shifted to the right which could represent n = 4 binding mode. This shift cannot be caused by weaker binding at the larger NaCl concentrations, because the experiments were performed at protein and nucleotide concentrations high enough to guarantee strong binding (vide supra). There is an analogy with the behaviour of the rotational correlation times observed for the complexation of gene-5 protein with M13 DNA (Fig. 6.5). In both cases steeper curves are observed at low ionic strength for the lower N/P ratios, which indicates a lower binding stoichiometry relative to the curves obtained at higher salt concentrations. At N/P ratios above the stoichiometric value, the rotational correlation time maintains a very high value as was also observed for the complexation with M13 DNA (see Fig. 6.5). For the n = 4 binding mode, as observed for M13 DNA, this can be expected in view of the large cooperativity factor. Whether the high $\phi$s observed for $(dT)_{25-30}$ beyond the stoichiometric N/P value arise from a transition from the n = 3 to the n = 4 binding mode as was observed for polynucleotides or the n = 3 mode itself has a large cooperativity, cannot be decided at this stage. We will return to this question below.

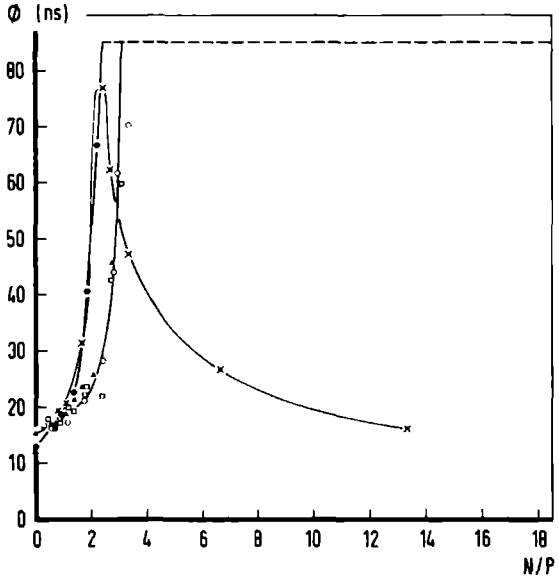



Figure 6 6 Rotational correlation times, $\phi$, of gene-5 protein / $(dT)_{25-30}$ mixtures as a function of N/P, obtained at 20 °C The sodium chloride and the nucleotide or the gene-5 protein concentrations were kept constant during the titration at less than 1 mM NaCl and 137 2 µM $(dT)_{25-30}$ (•), less than 1 mM NaCl and 75 0 µM gene-5 protein (•), 90 mM NaCl and 102 0 µM $(dT)_{25-30}$ (□ and ▲) or 135 mM NaCl and 75 0 µM gene-5 protein (o) The horizontal line at higher N/P ratios indicates high values of $\phi$ The exact values cannot be determined because of the fast-decaying protein fluorescence Also the curve obtained for the titration of 59 72 µM $(dT)_{18}$ at 1 mM NaCl and a constant gene-5 protein concentration is given (×)

### 6.4d Stopped flow dissociation kinetics

**Stopped flow salt jump experiments have been used to characterize the type of rate process and to determine the rate constant of the dissociation of gene-5 protein - oligonucleotide complexes. Only oligo(dT) complexes could be**

used; the smaller binding constant of oligo(dA)s makes the dissociation kinetics of oligo(dA) complexes very fast which implies that the reaction proceeds completely within the mixing time of about 4 ms, even at temperatures as low as 5 °C. At room temperature also the oligo(dT) complexes exhibited very fast reactions which proceeded largely within the mixing time. However, at a lower temperature, e.g. 8 °C, the dissociation of the oligo(dT) complexes could be followed without problems. Fig. 6.7 shows (as an example) the dissociation curve obtained after rapid mixing of a solution of the gene-5 protein - $(dT)_{12}$ complex with an equal volume of 2 M NaCl. The curve can be described by a single exponential, characterized by a relaxation time of 23 ms. For the other dissociation experiments, of which the results are collected in Table 6 II, the process can also be described by a mono-exponential function. This indicates that the dissociation rate only depends on the concentration of

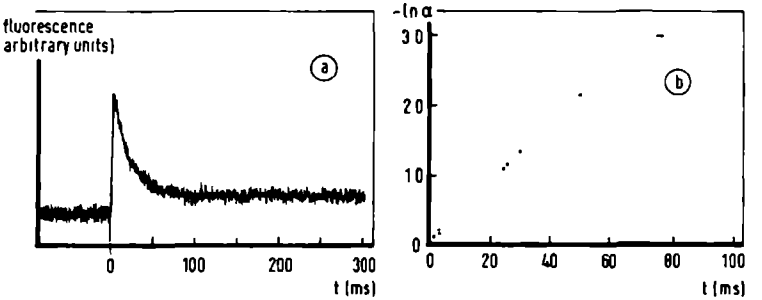



Figure 6 7 Dissociation kinetics of a gene-5 protein - $(dT)_{12}$ complex resulting from rapid mixing in a stopped flow apparatus of a solution containing 25 µM gene-5 protein and 75 µM $(dT)_{12}$ in a 10 mM sodium cacodylate buffer (pH 6 9) with an equal volume of a solution of 2 M NaCl in the same buffer, T = 8 °C a Time course of the dissociation as recorded directly from the display of the instrument The signal before mixing represents the fluorescence intensity after complete dissociation of the complex obtained in the previous run b Fit of the actual decay curve by means of a mono-exponential function which delivers a value of 23 ms for the relaxation time in this experiment

Table 6 II

Stopped flow irreversible dissociation experiments of gene-5 protein oligo(dT) complexes The sodium chloride concentration of the sample before and after the rapid mixing of the complex and salt solution save is indicated The experimental relaxation time $\tau$ is the average value obtained from a one-exponential fit of about 15 different experimental curves The uncertainty amounts from about 10 % for the lower values of $\tau$ to about 15 % for the higher values The ratio of total nucleotide to protein concentration is given by the parameter N/P

| nucleotide | [NaCl] (M) initial | final | temperature ($^{o}$C) | N/P | $\tau$(ms) |
|---|---|---|---|---|---|
| $(dT)_{12}$ | 0 | 2 00 | 8 | 3 | ≤ 8 |
| ,, | 0 | 1 50 | 8 | 3 | 13 |
| ,, | 0 | 1 00 | 8 | 3 | 21 |
| ,, | 0 | 0 50 | 8 | 3 | 45 |
| $(dT)_{18}$ | 0 | 2 00 | 8 | 3 | 70 |
| $(dT)_{25-30}$ | 0 | 2 00 | 8 | 3 | 100 |
| ,, | 0 | 2 00 | 8 | 3 | 86 |
| ,, | 0 | 2 00 | 13 | 3 | 37 |
| ,, | 0 | 2 00 | 16 | 3 | 28 |
| ,, | 0 115 | 2 05 | 8 | 4 | 76 |
| ,, | 0 25 | 2 13 | 8 | 4 | 99 |
| ,, | 0 30 | 2 15 | 8 | 4 | 85 |
| ,, | 0 | 2 00 | 8 | 6 | 49 |
| ,, | 0 | 2 00 | 8 | 8 | 52 |

bound protein and not on the distribution of the bound protein molecules over the nucleotide lattices. Furthermore, it is clear from Table 6 II that the experimental relaxation time $\tau$ increases with the chain length of the oligonucleotide. The corresponding decrease of the dissociation rate is in agreement

with the observed increase of the apparent binding constant with the chain length (see above). For $(dT)_{25-30}$ the dissociation process is slow enough to permit measurements at higher temperature. From the temperature dependency of $\tau$ thus obtained (see Table 6.II) an activation energy of 21 kcal per mole could be derived.

## 6.5 DISCUSSION

### 6.5a Dual mode binding of gene-5 protein to polynucleotides

The parameters characterizing the binding of gene-5 protein to polynucleotides in the so-called n = 4 mode have been determined for various ribo- and deoxyribopolynucleotides [37,70,chapter 3 of this thesis] and are well known at present. This binding mode prevails for a wide range of salt concentrations above 0.05 M NaCl, including the physiological ionic strength, and is characterised by a large cooperativity factor ($\omega \approx 500$). CD measurements conducted above 250 nm [70] revealed that at very low salt conditions also a binding mode with a stoichiometry of n = 3 is possible for poly(dA), poly d(I), poly d(A-T) and poly d(A,T). It was found as a second binding mode at higher protein concentrations. Upon titrating the polynucleotide with protein its occurence was demonstrated by a partial reversal of the CD changes found for the n = 4 mode. For poly(dT) and fd DNA the CD titration technique shows only the n = 3 binding mode at these low salt concentrations. However, as pointed out by Gray and coworkers, it is possible that the switch between the two modes during the titration of these DNAs is not detectable because it is not accompanied by a reversal of the CD change. Using the fluorescence titration method we as well observed both binding modes for poly(dA) and the viral DNA at very low salt concentration. The absence of the n = 3 mode at higher salt concentrations implies that the binding constant of this mode drastically decreases with increasing ionic strength. This means that it is much more sensitive to the ionic strength of the solution than the n = 4 mode. To further characterize the n = 3 binding, especially with respect to the magnitude of its cooperativity factor, we simulated the fluorescence titration curve for the simultaneous binding of a ligand in two different binding modes to an infinite lattice using the theory of Schwartz and Stankowski

(1979) [37]. First, the degree of saturation, $\alpha_i$, of the polynucleotide with protein bound in mode i was computed as a function of the P/N ratio. Subsequently, the fluorescence intensity, F, could be calculated from Eq. 6.1 with the concentration of protein bound in mode i, $[P_{b,i}] = [N_t]\ \alpha_i / n_i$, in which $[N_t]$ denotes the total nucleotide concentration in the sample and $n_i$ is the stoichiometry of the binding mode. For the calculation of the binding profiles in Fig. 6.1d and thereupon of the fluorescence titration curve in Fig. 6.1c, we used the following values for the binding parameters; stoichiometry: $n_1 = 4$, $n_2 = 3$; intrinsic binding constant: $K_{int,1} = 1.0 \times 10^{10}\ M^{-1}$, $K_{int,2} = 9.0 \times 10^{9}\ M^{-1}$; cooperativity factor: $\omega_{11} = \omega_{22} = 500$ and $\omega_{12} = \omega_{21} = 1$; total nucleotide concentration: $N_t = 10\ \mu M$; fluorescence quenching factor: $Q_{max,1} = 0.60$, $Q_{max,2} = 0.45$. The choice $\omega_{21} = \omega_{12} = 1$ means that proteins in both binding modes are mutually non-interacting. Comparison of the curves a and b with c in Fig. 6.1 reveals that the titration curve thus obtained is almost identical with the experimentally determined curves. Hence, the fluorescence titration curves can be described by a model that assumes binding in two different modes. Both modes have a large intramode cooperativity but are non-interacting and mutually exclusive. In addition calculations have been performed for a wide range of binding constants and cooperativity factors. The same curves are obtained if the variation of the values of $K_{int,1}$ and $K_{int,2}$ is constrained by the constant ratio $K_{int,1} / K_{int,2} = 1.11$, provided that the condition of strong binding is fulfilled. However, no agreement between the simulated and the experimental titration curves could be obtained when $\omega_{22}$ is small or when $\omega_{12}$ (= $\omega_{21}$) has a large value. Therefore we can conclude that also the n = 3 binding mode has a rather high intramode cooperativity and that an intermode cooperativity between the two stoichiometric ways of binding is absent. This absence of intermode cooperativity was also found by Kansey et al. from a comparison of measured and calculated CD titration curves [70]. In their caluculations these investigators adopted a ratio of $K_{int,1} / K_{int,2} = 1.25$, which is somewhat higher than our value of 1.11. The only consequence of this difference is that a maximum degree of saturation of 0.89 in the n = 4 mode at N/P = 0.241 is found while in our calculations this maximum is 0.68 at N/P = 0.223 (see Fig. 6.1d). Note that a maximum of 1.00 at N/P = 0.25 would indicate complete saturation of the polynucleotide with protein in the n = 4 binding mode.

### 6.5b Binding of gene-5 protein to oligonucleotides.

The binding of gene-5 protein to oligonucleotides differs in a number of aspects from the binding to polynucleotides, at least as far as binding to oligo(deoxyadenilic acid)s and oligo(deoxythymidilic acid)s is concerned. First, the stoichiometry of the binding is n = 3 (or perhaps somewhat smaller) up to a chain length of about 20 nucleotide residues, independent of the ionic strength. Secondly, there is a large difference between the cooperativity of oligonucleotide and polynucleotide binding. Binding to polynucleotides is always characterized by a high value of the cooperativity factor ($\omega \approx 500$, vide supra) for both possible binding modes. However, as we have found with various different techniques, it is evident that for the binding to oligonucleotides the existence of such a high cooperativity can be excluded. Thus, from the fluorescence depolarisation experiments it is obvious that the decrease of the rotational correlation time of the protein after the N/P ratio has increased above the stoichiometric value points to a redistribution of the bound protein molecules over the nucleotide lattices (see Fig. 6.6). This will only take place if the binding cooperativity is low compared with the configurational entropy factor, i.e. the number of ways a protein molecule can bind to a lattice. Furthermore, as expected for low binding cooperativity, the dissociation kinetics of gene-5 protein - oligo(dT) complexes, measured after addition of salt, follows a single exponential. This indicates that the dissociation rate is only dependent on the concentration of bound protein molecules irrespective as to whether these protein molecules are bound on isolated sites or next to each other. This dissociation behaviour is in sharp contrast to the dissociation kinetics of gene-5 protein - poly(dA) complexes for which a linear decay of the concentration of bound protein with time is observed (see Chapter 5). We have explained these dissociation kinetics by a mechanism in which the protein "slides" along the nucleotide chain from the ends of cooperatively bound protein clusters to isolated sites. This model also explains the dissociation of gene-5 protein from M13 DNA.

Unfortunately, it is not possible to determine quantitatively the cooperativity factor for the binding of the protein to oligonucleotides from the fluorescence experiments. Also the shape of the fluorescence titration curves does not allow a more precise estimation of $\omega$. For polynucleotides values of $\omega$ could be

estimated from binding isotherms or salt titration curves by means of fitting procedures based on the model of McGhee and Von Hippel for binding to an infinite lattice (see chapter 3). This model is not applicable to short nucleotide lattices. Nevertheless, the shape of the binding isotherms does confirm the occurrence of a low cooperativity for oligonucleotide binding because at salt concentrations at which the binding is weak a hyperbolic shape is observed. This is in sharp contrast to the sigmoidal shape obtained for the binding of gene-5 protein to polynucleotides under these conditions, which indicates a high cooperativity. The only number obtained for the cooperativity parameter of oligonucleotides is a value of about 5 found in our laboratory for the binding to oligo(deoxyadenylic acid)s [67]. This was deduced from the change in the chemical shift position of a gene-5 protein tyrosyl resonance upon titration of the protein with oligo(dA)s.

The third parameter which is important for the characterisation of the binding is the intrinsic binding constant, $K_{int}$. For the binding to oligonucleotides we have according to Epstein et al. [87] the following relationship between the binding constant $K_{int}$ and $\alpha$, the degree of saturation of an oligonucleotide with protein:

$$\alpha = (n/M) \sum_q \sum_j q\, \sigma(M,n,q,j)\, (K_{int}\,[P_f])^q \omega^j \;/\; \sum_q \sum_j \sigma(M,n,q,j)\, (K_{int}[P_f])^q \omega^j$$

$$(q = 0,1,2, \ldots \text{int}(M/n);\ j = 1,2, \ldots (q-1)) \qquad (6.6)$$

In this expression M is the number of nucleotide residues in the oligonucleotide, n is the size of the binding site (residues per protein), $[P_f]$ is the concentration of unbound protein and $\sigma(M,n,q,j)$ is the number of different ways in which q protein molecules can bind to a lattice of M residues and form j cooperative contacts. The configurational entropy factor $\sigma(M,n,q,j)$ is represented by

$$\sigma(M,n,q,j) = \binom{M-nq+1}{q-j}\binom{q-1}{j} \qquad (6.7)$$

In the titration of an oligonucleotide - gene-5 protein complex with salt, $K_{int}$ is not a constant parameter because it is the decrease of $K_{int}$ with increasing salt concentration which causes the lowering of $\alpha$, i.e. the dissociation of the complex. However, if the cooperativity parameter, $\omega$, is assumed to be constant (note that we found a very low cooperativity at all salt concentrations used) it follows from Eq. 6.6 that the product $K_{int}[P_f]$ is a constant for a particular value of $\alpha$, independent of the complex concentration. If for different complex concentrations salt titrations are performed, the salt concentration at which half saturation is achieved will be different and occur at different $[P_f]$ values. We are not able to derive $K_{int}$ directly from such experiments, but we can define a new binding constant, $K_{app}$ (see Eq. 6.3) which is linearly poproportional to the intrinsic binding constant $K_{int}$:

$$K_{app} = \rho(M,n,\omega)\ K_{int} \qquad (6.8)$$

The value of the proportionality constant $\rho(M,n,\omega)$ depends on the model used for the calculation of this constant and is also different for the various oligonucleotides. The advantages of this definition of the binding constant are that $K_{app}$ is independent of the binding model, that the free protein concentration at half saturation of the nucleotide can be directly obtained from fluorescence salt titrations without data fitting or the involvement of assumptions and that the salt dependency of $K_{int}$ and $K_{app}$ are the same. As discussed already this salt dependency follows from the determination of $K_{app}$ at different concentrations of the protein - oligonucleotide complex (see Fig. 6.4). Then according to Record's theory [73,74] the following relationship should exist between the binding constant and the salt concentration:

$$\log K_{app} = c \log [NaCl] + b \qquad \text{with } c = -(m'\psi + a) \qquad (6.9)$$

Here b and c are constants; b is the value for log $K_{app}$ at 1 M NaCl, $m'\psi$ is the number of sodium ions released from the DNA and a is the number of chloride ions released from the protein upon binding of a protein molecule. For binding to polynucleotides we have found previously already such a linear relationship with c = -4.0 ± 0.8 (see chapter 3). Eq. 6.9 is also applicable to the oligonucleotide binding and reveals an average value for c of -2.4 ± 0.2, independent of the type of oligonucleotide. So the same dependence of $K_{app}$ on the salt concentration is found for all oligonucleotides studied, which is however smaller than the salt concentration dependency of $K_{int}{}^{\omega}$ for polynucleotides.

Considering that the value of -4.0 for c holds for n = 4 polynucleotide binding, while the value of -2.4 from Fig. 6.4 applies to n = 3 oligonucleotide binding, this difference in c can be explained for the greater part by the release of a smaller number of sodium ions upon binding of gene-5 protein to oligonucleotides compared with the binding to polynucleotides in the n = 4 mode. Assuming that the number of sodium ions per phosphate bound to the uncomplexed DNA equals 0.78, both for polynucleotides as well as for oligonucleotides (in fact this value for $\psi$ holds for the binding to poly(rA), [73]) then $m'\psi$ = 3.1 (see Eq. 6.9) for the n = 4 polynucleotide binding (m' = 4) and $m'\psi$ = 2.3 for oligonucleotide binding (m' = 3). This means that upon complex formation with polynucleotides possibly also a chloride ion is expelled from the protein (see [37]), while for binding to oligonucleotides this is probably not so. Thus there is a close releationship between the decreased ionic strength dependency of the oligonucleotide binding and the smaller stoichiometry of this binding mode.

More insight in the ionic strength dependency of the binding of gene-5 protein to oligonucleotides can be obtained from the salt dependence of the relaxation time, $\tau$, which characterizes the dissociation of gene-5 protein - $(dT)_{12}$ complexes. From a linear plot of log $\tau$ versus the logarithm of the final salt concentration (not shown), using the values collected in Table 6.II, a slope of -1.5 ± 0.3 is obtained. This implies that d log $k_{diss}$ / d log [NaCl], in which $k_{diss}$ stands for the dissociation rate constant, equals 1.5 ± 0.3. Since the binding constant is directly proportional to $k_{diss}{}^{-1}$ or $\tau$, the slope of -1.5 ± 0.3 can be related with the parameter c from Eq. 6.9, which is equal to -2.4 ± 0.2. So the ionic strength dependency of the apparent

binding constant (which represents also the salt concentration dependency of the intrinsic binding constant) originates, within experimental error, to about the same extent from the dissociation and from the association reactions. The ionic strength dependency of the association rate is thus given by (d log $k_{ass}$ / d log [NaCl] = -0.9 ± 0 5).

### 6.5c The n = 3 modes do not represent the same binding in oligo- and polynucleotides

As has been discussed the binding of gene-5 protein to oligonucleotides as well as to polynucleotides (at very low salt concentration) is characterized by a stoichiometry of n = 3 or less and by a maximum quenching factor of the fluorescence of about 0.45. However the other binding properties are quite different. The polynucleotide binding appears to be highly cooperative and very sensitive to ionic strength, since no binding in this mode can be detected at salt concentrations above 50 mM NaCl. In contrast, oligonucleotide binding is nearly non-cooperative and shows a salt concentration dependency per residue which is comparable with that of the n = 4 polynucleotide binding mode. Therefore in our opinion the oligonucleotide and polynucleotide n = 3 modes represent a different type of binding.

From the sucrose gradient analysis of gene-5 protein [18] and gene-5 protein cross linking experiments [32] we know that the gene-5 protein binds to DNA as a dimer. The gene-5 protein binding has often been described as a system in which the two halves are bound to opposing DNA strands [17,33,34,44,65]. This idea is supported by the observation that in the crystal structure the protein forms a symmetric dimer with a twofold axis of symmetry [65]. However it is interesting to entertain the notion that both monomers of a dimer bind to the same DNA chain. Indeed we will demonstrate that oligonucleotides may prefer this state of complexation. To this end we simulated rotational correlation time versus N/P curves for a model in which the two halves of the dimer are bound to two different strands of $d(T)_{18}$ (model 1) and for a model in which the two halves of the dimer are bound to the same strand (model 2). In general, the fluorescence depolarisation for a (slowly) exchanging system in equilibrium is given by

$$A(t) = \Sigma_q A_q(0) f_q e^{(-t/\phi_q)} \qquad (6.10)$$

where $A_q(0)$ is the zero time anisotropy, $\phi_q$ is the rotational correlation time of complex q and $f_q$ is the fraction of the protein which is present in complex q. During the titration process, complexes of the type $S_2D$, $S_2D_2$, $S_2D_3$, $S_2D_4$, $S_2D_5$ and $S_2D_6$ (model 1) or of the type SD, $SD_2$ and $SD_3$ (model 2) can be formed between a strand S of 18 residues and the protein dimer D. Note that q = 0 for the free protein. For practical purposes Eq. 6.10 can be simplified to

$$A(t) = A(0) \Sigma_q f_q e^{(-t/\phi_q)} \qquad (6.11)$$

because it turned out (see chapter 4) that for the gene-5 protein / oligo(dT) systems the zero-time anistropy only varied to a limited amount (maximum total decrease was only 14 % throughout the titration). Theoretical curves for the rotational correlation time, as a function of the N/P ratio, can now be simulated. First, A(t) (Eq. 6.11) is calculated with the following input parameters. The fraction, $f_q$ , of the protein present in the particle of type q, is calculated from a combination of the law of mass conservation and expressions for the binding constants obtained from Epstein's theory [87] for the binding of a ligand to an oligonucleotide as a function of oligonucleotide concentration (see Appendix). Secondly, the rotational correlation time is assumed to be 10 ns for the free protein dimer and (2 + 10.q) ns for the various complexes. Combination of these different parameters allows a calculation of the anisotropy A(t) as a function of time (using Eq. 6.11) and as a function of nucleotide concentration. Subsequently, the effective rotational correlation time is calculated by fitting a mono-exponential function to the curve calculated for A(t). The values thus found for the effective rotational correlation time as a function of the nucleotide protein ratio have been plotted in Fig. 6.8a (model 1) and Fig. 6 8b (model 2). The same qualitative results are obtained if one introduces asymmetric binding in model 1 (the two DNA strands are allowed to be not completely matched but shifted in relation to each other) or an initiation step (the intrinsic binding constant for binding of a dimer to two empty strands has a smaller value then the intrinsic binding

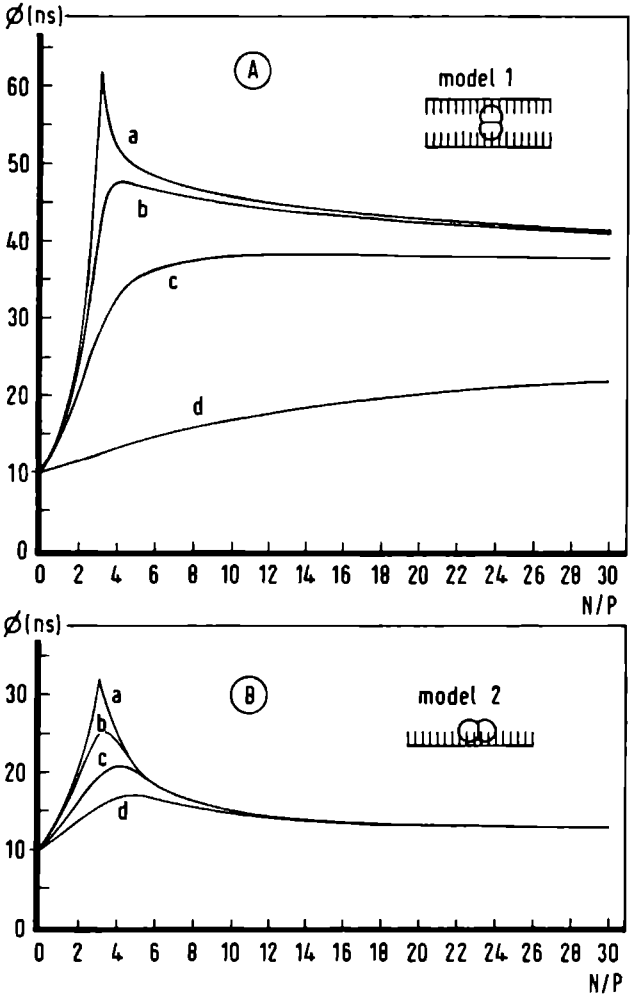

Ø(ns)
A
model 1
a
b
c
d
0
10
20
30
40
50
60
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30
N/P
Ø(ns)
B
model 2
a
b
c
d
0
10
20
30
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30
N/P

Figure 6 8 Rotational correlation times calculated for a single-strand DNA binding protein dimer as a function of N/P, the ratio mononucleotide / protein monomer concentration, for binding of the protein to a nucleotide lattice of 18 residues One protein monomer covers three nucleotides when bound to the DNA The protein monomer concentration is kept constant at 10 μM, the cooperativity parameter ω equals 1 and a time window of 7 ns is used (see text) The results for model 1 and model 2 are given in Fig A and B, respectively The $\phi$ values were obtained through a one-exponential fit to the simulated anisotropy decay curves The binding constant for the binding of a protein dimer was assumed to be $10^{12}$ (a), $10^{8}$ (b), $10^{7}$ (c) or $10^{4}$ (d) for the curves in Fig A and $10^{12}$ (a), $10^{8}$ (b), $10^{7}$ (c) or $10^{6}$ (d) $M^{-1}$ for the curves in Fig B

constant for binding of a dimer to a complex). A square window parameter of 7 ns was used. The window parameter is equal to the time interval in which the mono-exponential is fitted to the decay curve. In practice this is determined by the fluorescence lifetime. The shape of the depolarisation titration curves thus obtained is similair to the shape of the curves given previously for the binding to a nonanucleotide (chapter 4) and has been explained loc. cit.. The difference between the two models is most pronounced in the right hand part of the curves i.e. at an excess of nucleotide (high N/P ratios). With the second model (binding of a dimer along one strand) a fast decrease of the rotational correlation time is observed upon increasing the N/P ratio above the stoichiometric value (Fig. 6.8B). This is exactly what we have obtained experimentally (compare with Fig. 6.6). According to the first model (binding of a dimer between two different strands) the rotational correlation time maintains a rather high value (Fig. 6.8A) even at a large excess of nucleotide. In conclusion it is obvious that only model 2 can explain the course of the $\phi$ versus N/P curve of $(dT)_{18}$ complexes. Therefore in our opinion this model represents the binding of gene-5 protein to oligonucleotides, at least to oligo(deoxythymidilic acid)s. Note that the very high values obtained for the rotational correlation time of $(dT)_{25-30}$ complexes at excess nucleotide (see Fig. 6.6) can be explained by a transition from oligonucleotide to polynucleotide binding mode as has been observed also for $(dA)_{25-30}$ complexes [69]. Consistent with this would be a non-exponential decay in the dissociation experiments of $d(T)_{25-30}$ complexes at N/P ratios above the stoi-

chiometric point. However, the very low signal to noise ratio of dissociation time curves obtained at high N/P ratios prohibits experimental verification of this transition. For N/P values up to 8, an exponential dissociation time curve is observed (see Table 6.II). Thus up to this value the oligonucleotide binding mode still prevails.

6.5d Influence of the nucleotide chain length on the apparent binding constant and the effective dissociation rate of oligo(dT) - gene-5 protein complexes.

Comparison of the dissociation rates of $(dT)_{12}$, $(dT)_{18}$ and $(dT)_{25-30}$ - gene-5 protein complexes obtained under identical conditions, i.e. at an N/P ratio of 3 and a final salt concentration of 2 00 M, reveals that the values of $\tau$ increase with increasing chain length (see Table 6.II). This reflects a decrease of the dissociation rate which corresponds with an increase of the apparent binding constant. Actually, the chain length dependences of $\tau$ and $K_{app}$ are identical. If we assume that in solution the complex is mainly present in the form of single oligonucleotide chains to which protein dimers are bound in a non-cooperative way, then we would expect that both $\tau$ and $K_{app}$ are independent of the length of the oligonucleotide. Therefore we have to expand our model 2 of binding of a dimer along one nucleotide strand, which we will call henceforth the open complex, with the assumption that in solution there exists a strong tendency of these open complexes to dimerize by means of interactions between protein dimers bound to different strands. The complex formed by this dimerization will be called the closed complex. The dissociation of protein from the open complex is still considered as the only way to produce free nucleic acid and protein; moreover the rate of this reaction forms the rate determining step in the total dissociation process. However, it is clear that the concentration of the open complex is dependent on the value of the equilibrium constant between the closed and the open complex. Since the former is stabilized by contacts between bound protein dimers it can be expected that an increase of the chain length of the oligonucleotide at complete saturation with protein favoures the formation of the closed complex which enhances $K_{app}$ as well as $\tau$. Furthermore, upon lowering the initial degree of saturation of a certain oligonucleotide with protein the number of

protein contacts decreases which will have the same effect as a smaller chain length, namely a decrease of $\tau$. This is indeed what can be seen in Table 6.II for the complexes with $(dT)_{25-30}$ at 8 °C and a final salt concentration of about 2 M NaCl: upon increasing N/P ratios, i.e. upon lowering the degree of saturation, $\tau$ becomes smaller in general.

As far as we know now, the effects described here seem to be valid only for oligo(dT)s, because Alma et al. [67] have found that there exists no effect of the chain length on the average binding constant per ligand of oligo(dA) - gene-5 protein complexes. This followed from proton NMR titration studies. In addition we have observed very strong line broadening of proton resonances of the protein in NMR studies in which the protein was titrated with $(dT)_{10}(dA)$, indicating the formation of larger units. This extreme line broadening has never been found with $(dA)_8$ or other oligo(deoxyadenylic acid)s.

## 6.6 APPENDIX

In the theoretical calculation of the rotational correlation times the fraction $f_q$ of the protein present in complex q served as an input parameter (of Eq. 6.11). Below we indicate the derivation of these parameters. First consider the equilibrium for model 1

$$2\ S + q\ D \underset{\leftarrow}{\overset{K_q}{\rightarrow}} S_2D_q \qquad (q = 1,2, \ldots 6) \tag{6.12}$$

and for model 2

$$S + q\ D \underset{\leftarrow}{\overset{K_q}{\rightarrow}} SD_q \qquad (q = 1, 2, 3) \tag{6.13}$$

where D denotes the protein dimer which binds to the 18-mer S that was used as the lattice in our model calculations. It is assumed that the protein covers three nucleotides per monomer. To be able to calculate the concentrations of the complexes $S_2D_q$ and $SD_q$, we introduce the intrinsic binding constant $K_{int}$, which characterizes the complex formation of one protein to an isolated binding site on the lattice, and the cooperativity parameter, $\omega$, which defines the factor by which the binding constant increases when a protein binds adjacent to an already bound protein relative to isolated binding. The binding constants for the formation of the various complexes, as defined by Eqs. 6.12 and 6.13, can be found with the theory of Epstein [87]. For example in model 2 there are 13 ways to bind an isolated protein dimer to an 18-mer, due to potential binding site overlap. The binding constant $K_q$ can be given for both models by

$$K_q = \Sigma_j \sigma(M,n,q,j)\ K_{int}^{\,q}\ \omega^j \qquad (6.14)$$

in which j denotes the number of cooperative contacts. The number of ways q protein dimers can be bound by forming j cooperative contacts is given by the numbers $\sigma(18,6,q,j)$ in Table 6.III below. If we make use of the mass conservation and mass action law

$$[S_t] = [S] + 2\ \Sigma_q\ [S_2D_q] \qquad \text{(model 1)} \qquad (6.15)$$

$$[S_t] = [S] + \Sigma_q\ [SD_q] \qquad \text{(model 2)} \qquad (6.16)$$

$$[D_t] = [D] + \Sigma_q\ q.[S_2D_q] \qquad \text{(model 1)} \qquad (6.17)$$

Table 6 III

Entropy factors σ, which indicate the number of different ways a ligand (the protein dimer), binding six nucleotides per ligand, can bind to a oligonucleotide of 18 residues q is the number of ligands bound per oligonucleotide and j stands for the number of contacts between contiguous bound ligands In model 1 the ligand binds to opposing DNA strands, in which each dimer half is bound to a different strand In model 2 the dimer is bound to one DNA strand with the two dimer halves to adjacent parts of the nucleotide strand

| model 1 | | | model 2 | | |
|---|---|---|---|---|---|
| q | j | σ(18,6,q,j) | q | j | σ(18,6,q,j) |
| 1 | 0 | 16 | 1 | 0 | 13 |
| 2 | 0 | 78 | 2 | 0 | 21 |
| 2 | 1 | 13 | 2 | 1 | 7 |
| 3 | 0 | 120 | 3 | 2 | 1 |
| 3 | 1 | 90 | | | |
| 3 | 2 | 10 | | | |
| 4 | 0 | 35 | | | |
| 4 | 1 | 105 | | | |
| 4 | 2 | 63 | | | |
| 4 | 3 | 7 | | | |
| 5 | 1 | 4 | | | |
| 5 | 2 | 24 | | | |
| 5 | 3 | 24 | | | |
| 5 | 4 | 4 | | | |
| 6 | 5 | 1 | | | |

$$[D_t] = [D] + \Sigma_q \, q.[SD_q] \qquad \text{(model 2)} \qquad (6.18)$$

and

$$[S_2D_q] = K_q \, [S]^2 \, [D]^q \qquad \text{(model 1)} \qquad (6.19)$$

$$[SD_q] = K_q \, [S] \, [D]^q \qquad \text{(model 2)} \qquad (6.20)$$

in which $[S_t]$ and $[D_t]$ are the total nucleotide and protein-dimer concentrations present in the sample, it is easy to demonstrate that for model 1

$$[D_t] = [D] + [S]^2 \; \Sigma_q \, ( q \, K_q \, [D]^q) \qquad (6.21)$$

in which

$$[S] = \{ -1 + (1 + 8 \, [S_t] \, \Sigma_q \, (K_q \, [D]^q))^{1/2} \} \; / \; \{4 \, \Sigma_q \, (K_q \, D^q)\} \qquad (6.22)$$

and for model 2

$$[D_t] = [D] + [S_t] \, \Sigma_q \, (q \, K_q \, [D]^q) \; / \; ( \, 1 + \Sigma_q \, (K_q [D]^q)) \qquad (6.23)$$

$$[S] = [S_t] / (1 + \Sigma_q K_q [D]^q) \qquad (6.24)$$

For a given value of $[D_t]$, $[S_t]$ and $K_q$, with the aid of the equations given, the free protein dimer concentration [D] and subsequently [S] and $[S_2D_q]$ or $[SD_q]$ can be calculated which in turn yield the $f_q$ values. The $f_q$ values so obtained were used in Eq. 6.11.

CHAPTER 7

# A TWO DIMENSIONAL NMR STUDY OF THE BACTERIOPHAGE M13 GENE-5 PROTEIN - HEXADEOXYADENYLIC ACID COMPLEX

## 7.1 ABSTRACT

A two dimensional nuclear magnetic resonance (NMR) study was performed of hexadeoxyadenylic acid and of the complex of this molecule with the gene-5 protein encoded by bacteriophage M13. In the absence of the gene-5 protein the connectivity patterns characteristic of right handed helices were observed in the 2D NOE spectra of the oligonucleotide. In the 2D NOE spectra of the complex changes in these connectivity patterns were observed, relative to the patterns obtained for the free hexanucleotide. The results are interpreted by considering a rotation of the terminal bases of the nucleotide to a syn-type conformation. A prerequisite for this conclusion is that the patterns obtained for the hexanucleotide in 2D NOE spectra of the complex can unambiguously be correlated to the structure of the hexanucleotide in the complexed form. It will be shown that this condition is fulfilled.

## 7.2 INTRODUCTION

The gene-5 protein encoded by bacteriophage M13 has been the subject of many nucleic acid binding studies [see section 1.3 for references; Chapters 3-6 of this thesis]. The protein can bind in two modes: to polynucleotides the gene-5 protein binds in a stoichiometric ratio of four nucleotides per protein molecule, while to smaller nucleotides (less than about 25 residues) or to polynucleotides at very low ionic strength three nucleotides are covered per protein. In both binding modes there is an electrostatic contribution to the binding, resulting from interactions between the phosphates of the nucleotide backbone and arginyl or lysyl residues. Tyrosyl and phenylalanyl residues

are thought to be involved in non-electrostatic stacking interactions with the nucleic acid bases. Experimental data sofar available indicate that the protein secondary structure is not influenced by the binding.

Not much is known about the structural changes of nucleic acids upon binding to the gene-5 protein. Circular dichroism changes at wavelengths above 250 nm observed for single stranded nucleic acids upon binding to the gene-5 protein can be interpreted to result from an uncomplete destacking of the bases [147,70]. The combination of chemical shifts of adenyl resonances and nuclear Overhauser effects in proton nuclear magnetic resonance binding studies of gene-5 protein - oligo(dA) complexes appeared to be consistent with a conformation in which aromatic residues are stacked upon adenyl bases in the vicinity of the base H8 proton, located relatively far from the base H2 proton and in which the glycosydic torsion angle is about $0^{o}$; the sugar proton that is closest to the aromatic protein residues is the H1' proton [69].

In order to obtain more insight regarding structural changes which might occur in nucleic acids upon binding to gene-5 protein we performed two-dimensional proton NMR experiments. For practical reasons we have chosen the hexanucleotide $d(A)_6$ as a probe to study the complexation. The adenyl residues provide a tool to study the important protein-base stacking interactions. A comparison between the 2D NMR spectra of $d(A)_6$ alone and of $d(A)_6$ complexed to gene-5 protein shows that large changes in the glycosidic bond angle may occur.

## 7.3 MATERIALS AND METHODS

### 7.3a Materials

Gene-5 protein was isolated and purified as described previously [60], except that during sonication of Escherichia coli no DNAse I was added. Concentrations of the gene-5 protein were measured using a molar absorption coefficient of 7100 $M^{-1}$ $cm^{-1}$ at the absorption maximum (276 nm) [29]. For NMR measurements the protein solution was concentrated by ultrafiltration and lyophilized two times from 99.75 % $D_2O$.

The hexanucleotide $d(A)_6$ was synthesized using the phosphotriester method [148]. It lacks the 5' and 3' terminal phosphates and was used in the $Na^+$ form. The molar absorption coefficient (257 nm) of 10320 $M^{-1}$ $cm^{-1}$ was estimated from the literature [137]. All NMR samples contained a buffer of 1 mM sodium cacodylate at pH 6.9. The salt concentrations of the samples are indicated in the figure legends.

### 7.3b NMR spectra

Proton NMR experiments were performed on a Bruker 300 MHz or a Bruker 500 MHz spectrometer, operating in the Fourier transform mode. Suppression of the water signal was performed by irradiation. Typical values of the irradiation power were 30 dB below 5 Watt. In all experiments the carrier frequency was placed in the middle of the spectrum. One-dimensional NMR spectra are presented in the absorption mode and two-dimensional spectra in the absolute value mode. Chemical shifts are quoted relative to 3-trimethylsilyl-(2,2,3,3-D) propionate (abbreviated as TSP); downfield shifts are defined positive.

Experiments concerning two dimensional correlated spectroscopy (COSY) [149-151] and two-dimensional nuclear Overhauser spectroscopy (NOESY) [152,153] were performed independently or combined in a COCONOSY experiment [154]. Pulse phase rotations and Fourier transformations were performed in such a way to obtain absolute value spectra. In this paper only the spectra are shown which are obtained at a field of 500 MHz. One dimensional spectra were not resolution enhanced prior to Fourier transformation. Two dimensional spectra have been resolution enhanced by using a non-phase shifted sine-filter for both frequency directions. Specific experimental conditions are given in the figure legends.

## 7.4a The $d(A)_6$ proton spectrum

The 500 MHz $^1H$ NMR spectrum of $d(A)_6$ recorded at 5 $^oC$ is shown in Fig. 7.1. From low to high field the resonances are found in the following order. the base H8 and H2 protons resonate between 7.4 and 8.2 ppm (Fig. 7.1B), the H1' protons between 5.4 and 6.1 ppm (Fig. 7.1C), the H3', H4' and H5'/H5" protons between 3.5 and 5.3 ppm (Fig. 7.1A) and the H2'/H2" protons between 2.0 and 2.6 ppm (Fig. 7.1D). As observed for other oligomer DNA systems the H2'/H2" spectral region is a rather complex cluster of resonances involving multiple J-couplings. The isolated cluster of resonances around 2.1 ppm belongs to the 5' terminal H2' proton [69,155]. The six H1' protons give rise to four resonances (Fig. 7.1C) with relative intensities of 1:1:1:3.

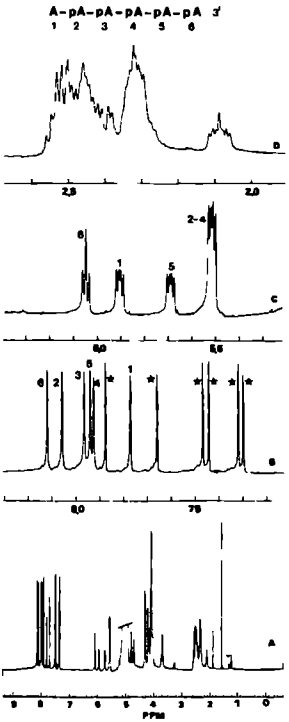
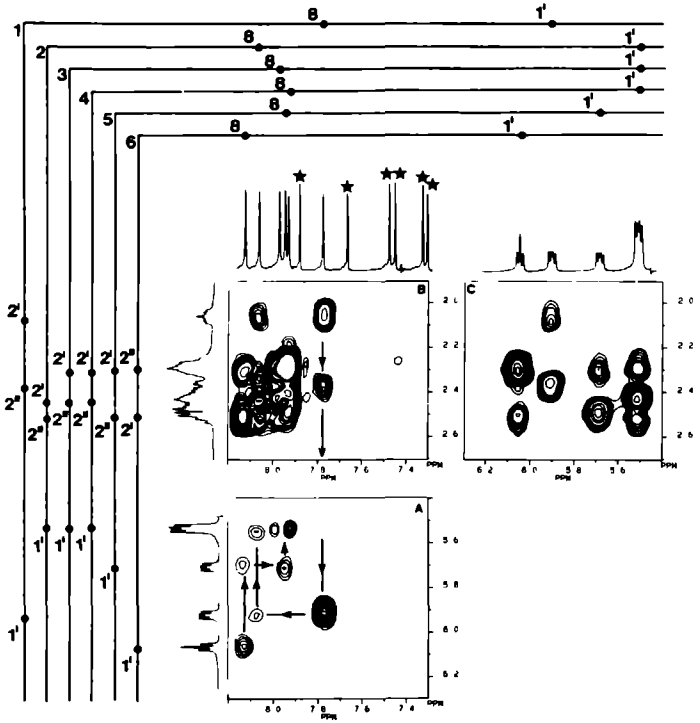



Fig 7 1 500 MHz $^1H$ NMR spectrum of $d(A)_6$, dissolved in $D_2O$, recorded at 5 $^oC$ The sample contained 22 2 mM $d(A)_6$ and 250 mM NaCl and the pH was 6 9 116 transients were recorded using quadrature detection stored in a 16 K memory No resolution enhancement procedures were applied A Total spectrum, the open triangles indicate resonances decended from contaminations in the sample B Subspectrum exhibiting H8 and H2 resonances The starred resonances arise from H2 protons The numbers added to the other resonances which arise from H8 protons, refer to the residues in the $d(A)_6$ sequence C Subspectrum exhibiting H1' resonances The numbers refer to the residues in the $d(A)_6$ sequence D Subspectrum exhibiting H2' / H2" resonances

The H8 and H2 resonances (Fig. 7.1B) can be distinguished by making use of the very different relaxation properties of the H8 and H2 spins. The longitudinal relaxation time of the H2 spins is much longer than that of the H8 spin and measurement of this relaxation time permitted the discrimination of the H2 and H8 resonances. Further analysis of the spectrum of $d(A)_6$ was achieved by application of sequential assignment procedures developed for double helical DNA fragments [156-158]. Intra- and inter- residue connectivities of resonances in NOESY spectra between a base H8 proton and sugar H1' or H2'/H2" protons allow the assignment of these resonances to specific residues in the DNA sequence. However, at ambient temperatures i.e. at 22 $^{o}$C and 27 $^{o}$C NOE connectivities could not be observed in NOESY spectra recorded at 300 MHz and 500 MHz respectively. Only by decreasing the temperature of the sample to 5 $^{o}$C, weak NOE connectivities appeared in the spectrum. Cross peaks observed between H8 and H1', H8 and H2'/H2" resonances and between H1' and H2'/H2" resonances are shown in expanded regions of the NOESY spectrum in Fig. 7.2. Examination of the connectivities between the H8 resonances and the sugar H1' resonances (Fig. 7.2A) shows that the standard sequential assignments can only be performed to a limited extent because of the overlap of three of the H1' resonances. The positions of the H8 resonances of the terminal and penultimate bases can be established; the connectivity pathways are indicated in Fig. 7.2A. Assignments for cross peaks between the H1' resonances of the central residues 3 and 4 and their H8 resonances could not be ascertained unambiguously at this stage. However, combination of the information contained in the connectivities between H1' and H2'/H2" resonances in the NOESY spectrum (Fig. 7.2B and C) and in the COSY spectrum (Fig. 7.3) yields sufficient additional data to be able to complete the assignment of the H8 resonances of the bases 3 and 4. From the intensity of the NOE crosspeaks in Fig. 7.2C and from the coupling patterns of the cross peaks in the COSY spectrum (Fig. 7.3) it can be concluded that for the 3' terminal residue the H2' resonance is at low field with respect to the H2" resonance. The assignments deduced in this way are collected in Table 7.1.


1
2
3
4
5
6
8
1'
2'
2"
A
B
C
ppm

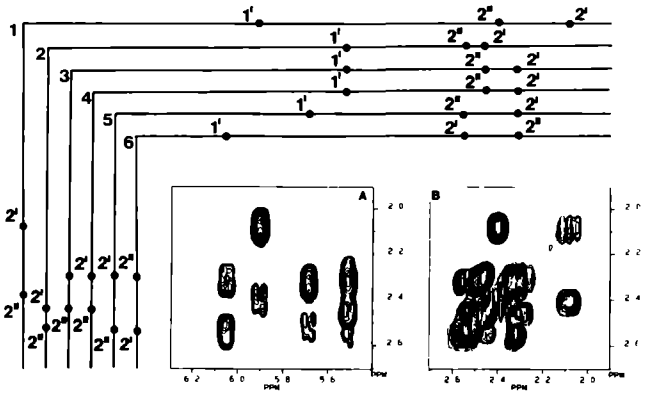

1
2
3
4
5
6
1'
2'
2"
A
B
ppm

Fig 7 2 Parts of the 500 MHz NOESY spectrum of $d(A)_6$ derived from a COCONOSY experiment performed at 5 °C The sample contained 22 2 mM $d(A)_6$ and 250 mM NaCl, pH = 6 9 The data set consists of 2K data points in the $\omega_2$-direction and 256 data points in the $\omega_1$-direction, 112 FIDs were recorded for each experiment The spectral width was 5000 Hz and quadrature detection was used with the carrier frequency in the middle of the spectrum A mixing time of 500 ms was used Absolute value contour plots are given Subspectrum A (spectral region $\omega_1$ = 5 4 to 6 3 ppm, $\omega_2$ = 7 3 to 8 2 ppm) exhibits connectivities between H1' and H8 resonances, subspectrum B (spectral region $\omega_1$ = 1 9 to 2 7 ppm, $\omega_2$ = 7 3 to 8 2 ppm) exhibits connectivities between H2' / H2" and H8 resonances, subspectrum C (spectral region $\omega_1$ = 1 9 to 2 7 ppm, $\omega_2$ = 5 4 to 6 3 ppm) exhibits connectivities between H2' / H2" and H1' resonances Resonance assignments of the H8, H1' and H2' / H2" resonances of the different adenyl residues are indicated above and at the left hand side of the contour plots The numbering of the corresponding adenyl residues is indicated in the upper left hand corner

Fig 7 3 Parts of the 500 MHz COSY spectrum of $d(A)_6$ derived from a COCONOSY experiment performed at 5 °C For the relevant parameters see the caption of Fig 7 2 The spectral region A ($\omega_1$ = 1 9 to 2 7 ppm, $\omega_2$ = 5 4 to 6 3 ppm) exhibits J-connectivities between H1' and H2' / H2" resonances and the spectral region B ($\omega_1$ = 1 9 to 2 7 ppm, $\omega_2$ = 1 9 to 2 7 ppm) exhibits J-connectivities between H2' and H2" resonances Above the contour plots the assignments of the H1', H2' and H2" resonances arising from the different adenyl residues number 1 through 6 (upper left corner) are indicated The numbering of the adenyl residues corresponds with that indicated in Fig 7 1

Table 7 I
Resonance positions (in ppm) of H8, H1', H2' and H2" resonances of $d(A)_6$ at 5 °C

| residue number | resonance | | | |
|---|---|---|---|---|
| | H8 | H1' | H2' | H2" |
| 1 | 7 78 | 5 92 | 2 08 | 2 38 |
| 2 | 8 07 | 5 52 | 2 42 | 2 55 |
| 3 | 7 97 | 5 52 | 2 30 | 2 43 |
| 4 | 7 93 | 5 52 | 2 30 | 2 43 |
| 5 | 7 95 | 5 69 | 2 31 | 2 51 |
| 6 | 8 14 | 6 06 | 2 50 | 2 31 |

### 7.4b Complex formation between $d(A)_6$ and the gene-5 protein

Binding of $d(A)_6$ to the gene-5 protein was studied by adding aliquots of a concentrated $d(A)_6$ solution to a gene-5 protein solution, see Fig. 7.4. It followed from these experiments that three nucleotides are covered by one protein monomer, in agreement with earlier results from our laboratory [67].

Subsequently, reversed experiments were performed in which the protein was titrated to a $d(A)_6$ solution, see Fig. 7.5. In combination with the assignments discussed in the preceding section, the positions of the H8, H2 and H1' resonances of the DNA hexamer were obtained from these experiments as a function of the protein nucleotide ratio. The results are collected in Fig. 7.6. The H8 and H1' resonances are clearly more broadened than the H2 resonances. Examination of the changes in chemical shift as a function of the protein nucleotide ratio shows that an upfield or downfield shift of an H1' resonance is accompanied by a shift of the intraresidue H8 resonance in the same direction.

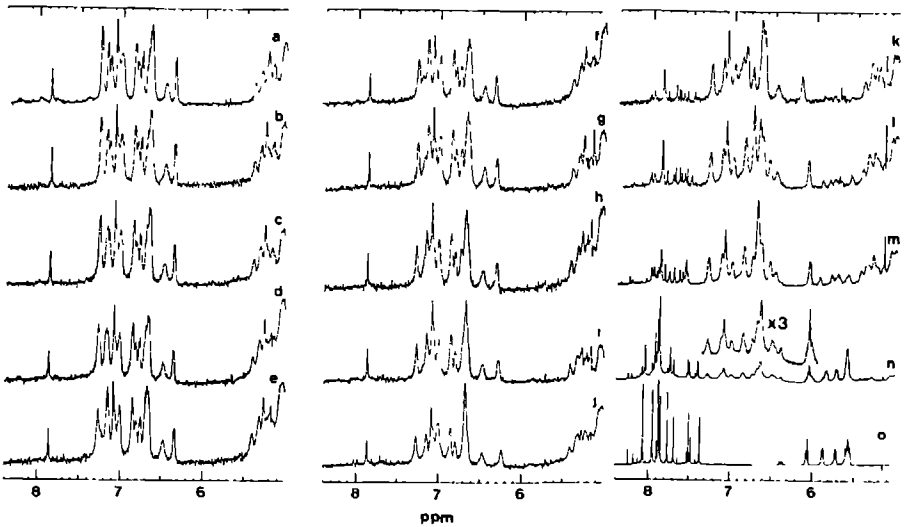



Fig. 7.4. Titration of a 0.35 mM gene-5 protein solution with $d(A)_6$ at 32 °C. The subspectrum between 5.0 and 8.5 ppm is shown, in which the protein aromatic protons resonate. Curve a shows the spectrum of pure gene-5 protein and curve o shows the spectrum of pure $d(A)_6$. From curve b to curve n the nucleotide concentration gradually increases, up to the point of excess nucleotide (curve n). The mononucleotide to protein ratio, N/P, for the several curves is a: 0, b: 0.10, c: 0.33, d: 0.43, e: 0.65, f: 0.86, g: 1.08, h: 1.30, i: 1.51, j: 1.73, k: 1.95, l: 2.16, l: 2.38 and n: 9.94.

**It is further noted that for all nucleotide protein ratios studied the exchange is fast on the NMR time scale. Under these conditions it is in principle possible to gain information on the structure of the $d(A)_6$ molecule bound to the protein by the observation of NOE effects between the resonances of the hexamer when present in excess to the protein [159]. Then knowledge of the hexanucleotide proton chemical shifts as a function of the protein and nucleotide concentrations is very helpful. It permits the choice of conditions where the dispersion of the resonance positions is optimal, so that in recording NOESY spectra the overlap between expected NOE cross peaks can be avoided as far as possible. The protein / mononucleotide ratios, P / N, for which 2D**

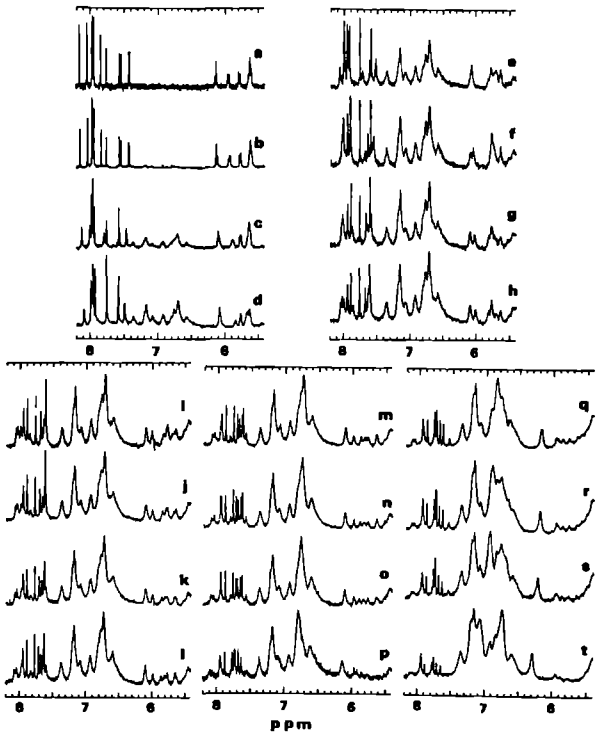
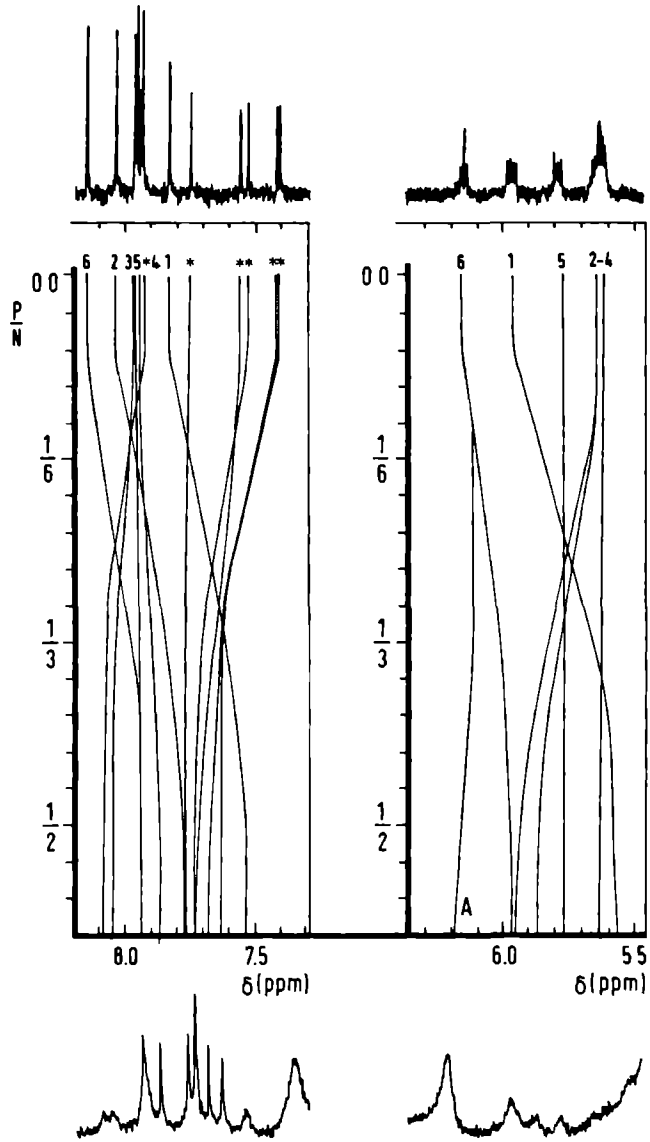



Fig. 7.5. Titration of a 2.21 mM $d(A)_6$ solution with gene-5 protein at 27 °C. The subspectrum between 5.4 and 8.2 ppm is shown, in which the H1',H8 and H2 protons of the hexanucleotide and the aromatic protons of the protein resonate. The protein to mononucleotide ratio P / N is a: 0, b: 0.078, c: 0.116, d: 0.115, e: 0.194, f: 0.223, g: 0.271, h: 0.291, i: 0.310, j: 0.330, k: 0.349, l: 0.369, m: 0.388, n: 0.407, o: 0.427, p: 0.491, q: 0.556, r: 0.621, s: 0.686 and t:0.946.

Fig. 7.6 (See opposite page). Chemical shift of the H8 and H2 resonances (left) and H1' resonances (right) of $d(A)_6$ as a function of the ratio of the number of protein monomers to nucleotides (P/N). The data were derived from the titration of 2.2 mM $d(A)_6$ with gene-5 protein at 27 °C. Peak numbers refer to the $d(A)_6$ residue number, see Fig. 7.1. Peak A is a gene-5 protein tyrosine resonance. At the top $d(A)_6$ subspectra are shown in the absence of protein and at the bottom $d(A)_6$ subspectra are given at a gene-5 protein / $d(A)_6$ ratio of 0.621. At high P/N ratios two nucleotide resonances (H8 peaks from the residues 5 and 6) and one sharp protein resonance (from a histidyl proton) coincide at 7.92 ppm.


6
2 3 5 *4 1
*
**
**
P/N
0 0
1/6
1/3
1/2
8.0
7.5
δ(ppm)
6
1
5
2-4
A
6.0
5 5
δ(ppm)

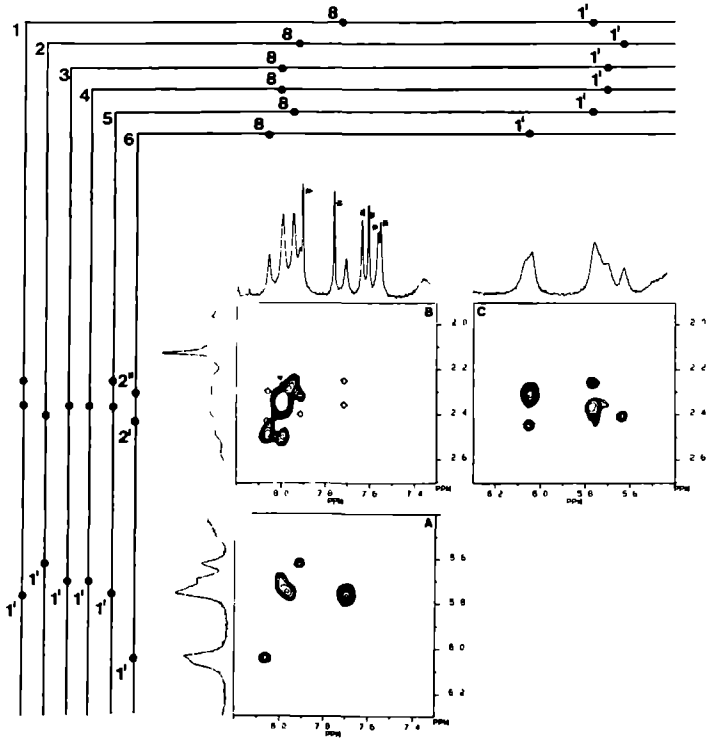



Fig. 7.7. Parts of the 500 MHz NOESY spectrum recorded for a mixture of *gene-5 protein and* $d(A)_6$ *in a ratio of protein to mononucleotide P/N = 0.176* at a temperature of 27 °C. The sample contained 7.55 mM $d(A)_6$, 1 33 mM gene-5 protein and 50 mM NaCl, at pH 6.9. The data set consists of 2K data points in the $\omega_2$-direction and 256 data points in the $\omega_1$-direction; 112 FIDs were recorded for each experiment. The spectral width was 5000 Hz and quadrature detection was used with the carrier frequency in the middle of the spectrum. A mixing time of 50 ms was used. Absolute value contour plots are given. The spectral regions exhibited correspond with those presented in Fig. 7.2. The

numbers in the upper left hand corner refer to the numbers of the residues of $d(A)_6$ presented in Fig 7 1 The resonance assignments of the H8, H1' and H2'/H2" resonances of the different adenyl residues are indicated above and at the left hand side of the contour plots The starred resonances above contour plot B arise from the adenine H2 protons The open diamonds in contour plot B indicate the position of the cross peaks expected from H8 / H2' and (or) H8 / H2" spins The cross peaks observed in spectral region B most likely arise from connectivities between adenyl H8 and protein protons

NOE spectra were recorded are 0 117 and 0.176 (compare Fig. 7.6); one of the results is presented in Fig. 7.7 (P / N = 0.176), in which expanded regions of the NOESY spectrum are given. Comparison with the patterns obtained for the pure $d(A)_6$ (Fig. 7 2) shows that the positions of the cross peaks have altered, as expected from Fig. 7.6. There is a striking difference with respect to cross peak intensities; as mentioned in the preceding section the 2D NOE spectra of the pure hexamer were recorded at 5 $^{o}$C, while the spectra of the mixture were obtained at 25 $^{o}$C, where no cross peaks are observed for the pure compound. Thus, the effective correlation time of the DNA hexamer is increased to such an extent, as a result of binding to the protein, that intramolecular NOEs become visible. In the spectral region where connectivities between H8 and H1' resonances are observed (Fig. 7.7A) only one cross peak per H8 resonance is observed (in contrast to Fig. 7.2); these cross peaks are interpreted to arise from intraresidue connectivities. Further examination of the results presented in Figs. 7.7B and 7.7C shows that a number of connectivities observed in the H8 - H2'/H2" spectral region (Fig. 7.7B) do not fall at the expected resonance positions of the H2'/H2" protons which can be derived from the H1' - H2'/H2" spectral region (Fig. 7.7C). This indicates that the connectivities observed in Fig. 7.7B arise from magnetization transfer between the H8 spins of the DNA and the methylene or methyl proton spins from the protein.

## 7.5 DISCUSSION

It is well-known that the adenyl bases in single stranded deoxyadenylic acids have a strong tendency for stacking. This leads to the formation of right handed single stranded helices in solution [160]. In the present paper these conclusions are reinforced for $d(A)_6$ on the basis of NOEs observed between base H8 and sugar proton resonances. Connectivity patterns characteristic for right handed helices are observed for this oligonucleotide in the 2D NOE spectra (cf. Fig. 7.2) [156-158]. It is interesting to note that for the 3' adenyl residue the order of the positions of the 2' and 2" sugar proton resonances is reversed with respect to that observed in regular B-type helices. Perusal of the literature indicates that this is often observed for 3' terminal residues, even in double helical molecules. The mentioned NOE connectivity patterns were observed at 5 $^{o}$C. At ambient temperatures it is no longer possible to detect NOE cross peaks for $d(A)_6$. This has most likely to do with rapid stacking - destacking processes, which take place at these temperatures.

As has been mentioned in the preceding section, complex formation between $d(A)_6$ and gene-5 protein drastically changes this picture. Now NOEs can be observed at ambient temperatures with a much better signal to noise ratio than for the free oligonucleotide at low temperature, even if the DNA is present in significant excess with respect to the protein (Fig. 7.2). As has been mentioned this can be attributed to the increase of the effective correlation

Fig 7 8 Cross sections along the $\omega_1$-direction of the NOESY spectrum of $d(A)_6$ (A, D, G), the NOESY spectrum of the gene-5 protein - $d(A)_6$ complex at a P / N ratio of 0 117 (B, E, H) and the NOESY spectrum of the gene-5 protein - $d(A)_6$ complex at a P / N ratio of 0 176 (C, F and I) For the relevant parameters see the captions of the Figs 7 2 and 7 7 The cross sections are taken at the $\omega_2$ resonance positions of the H8 protons of residue 6 (A, B, C), of residue 2 (D, E, F) and of residue 1 (G, H and I) The peaks marked with stripes in the cross sections C and F arise from H8 / protein proton cross relaxation

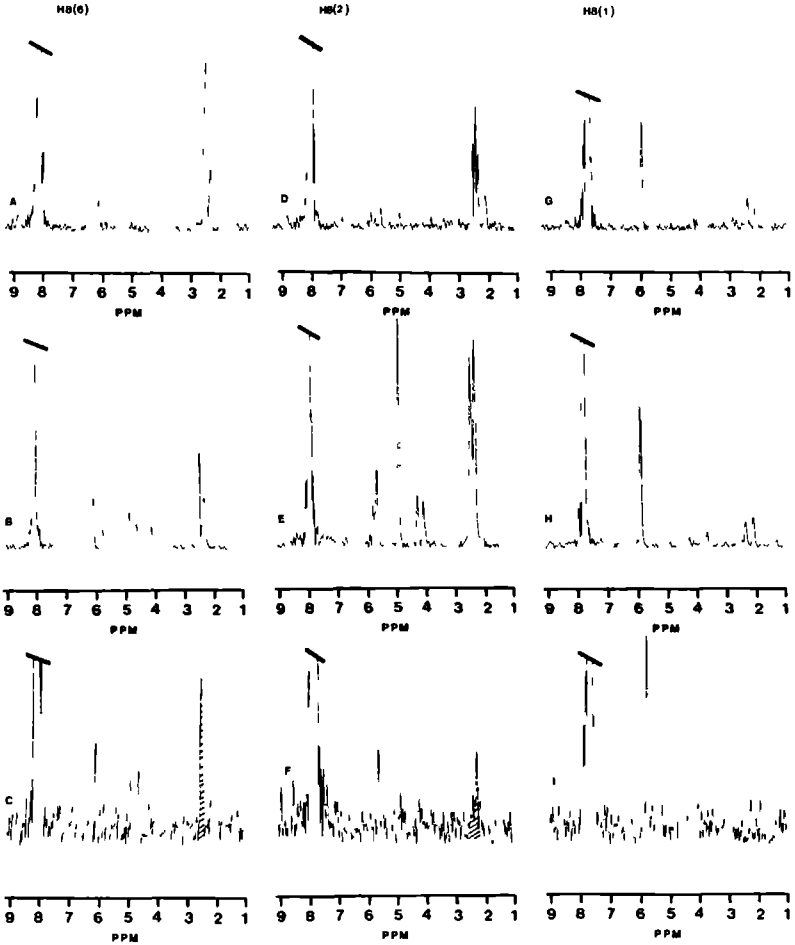
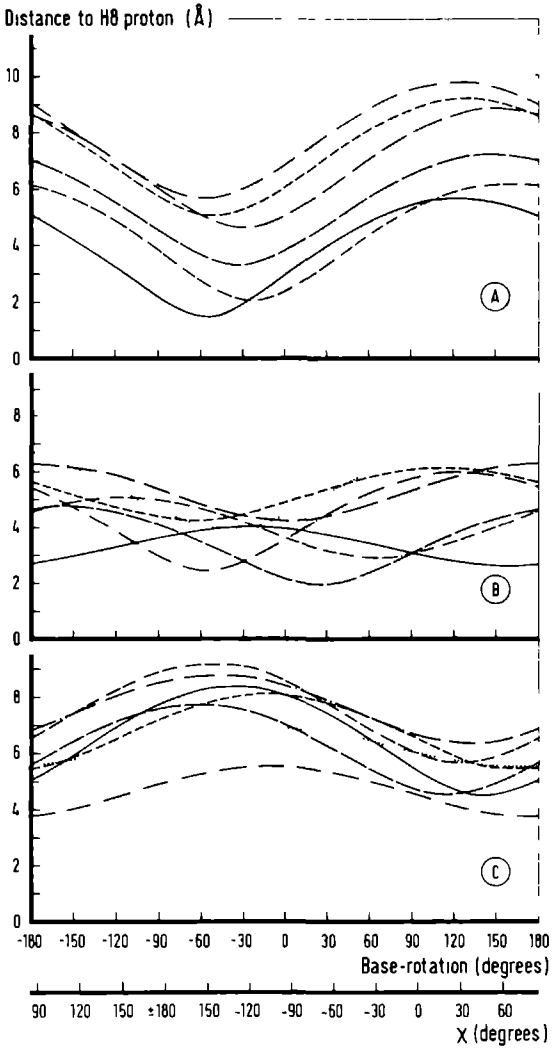

H8(6)
H8(2)
H8(1)
A
B
C
D
E
F
G
H
I
9 8 7 6 5 4 3 2 1
PPM

Fig. 7.9. Distances between sugar-protons in adenylic acid molecules, having a B-DNA type structure, and the base H8 protons as a function of the base rotation around the glycosidic bond. A rotation of zero degrees represents the structure of B-DNA. The relation between the base rotation and the glycosidic torsion angle x is also given. x is defined by the angle ∠(O4'-C1'-N9-C4). Distances of a H8 proton to 5' neighbour sugar protons (A), to the intraresidual sugar protons (B) and to the 3' neigbour sugar protons (C) are given. The different curves stand for
H1' (———), H2' (— — —), H2" (·— ·— ·—), H3' (—·—·—),
H4' (— — —), H5' (·— ·— ·—) and H5" (·····).

time of the small molecule. The theory developed by Clore and Gronenborn [159] to describe the influence of this phenomenon in 1D spectra is extended in the Appendix (section 7.6) for 2D spectra. A model system is considered that consists of a solution of small molecules with two spins (with a short rotational correlation time) and large molecules (with a long rotational correlation time) which can form complexes. When the small molecule exchanges between the free state and the bound state cross peaks between the resonances of the small molecule may be observed which reflect the structure of this molecule in the bound state. Accordingly, the cross peaks observed for $d(A)_6$ in the hexanucleotide - protein solution are taken to reflect the structure of the hexanucleotide in the complex.

To obtain further insight into the changes which occur upon complex formation between $d(A)_6$ and gene-5 protein we took cross sections along the $\omega_1$ direction in NOESY spectra taken of three different samples: pure $d(A)_6$ (see Fig. 7.2), P / N = 0.117 (NOESY spectrum not shown) and P / N = 0.176 (see Fig. 7.7). These cross sections correspond with the resonances of H8 protons of residues 1, 2 and 6. Examination of the cross sections (see Fig. 7.8) shows that the relative intensity of the H8 / H1' cross peaks is increasing at the expense of the intensity of the cross peaks between H8 / H2', H2" for increasing protein / $d(A)_6$ ratios. At the highest ratio the intensity of


Distance to H8 proton (Å)
10
8
6
4
2
0
A
8
6
4
2
0
B
8
6
4
2
0
C
-180 -150 -120 -90 -60 -30 0 30 60 90 120 150 180
Base-rotation (degrees)
90 120 150 ±180 150 -120 -90 -60 -30 0 30 60
χ (degrees)

the H8 / H2',H2" cross peaks has even reduced to zero. The peaks that are still seen at $\omega_1$ = 2.2 to 2.6 ppm arise from H8 / protein cross relaxation, in accord with the one-dimensional NOE experiments performed earlier for the gene-5 protein - $d(A)_8$ complex [69].

In addition to the relative intensity increase of H8 / H1' cross peaks, arising at increasing protein / $d(A)_6$ ratios (Fig. 7.8), a disappearance of interresidual H8 / H1' cross peaks is observed (compare Figs. 7.2 and 7.7). The most straightforward explanation of these changes is that the glycosidic torsional angle alters its value upon complex formation of $d(A)_6$ with the protein. To illustrate this, the distance of an H8 base proton to the intrasidual sugar ring protons and to the sugar protons of the 3' and 5' - neighbouring sugar was calculated as a function of the torsional angle $\chi$ (c.f. Fig. 7.9). As the starting structure an idealized type helix [161] was used. Examination of Fig. 7.9 shows that a base rotation of $150^{o} \pm 30^{o}$ (see upper horizontal scale in Fig. 7.9) leads to increased distances between the H8 base proton and the 5' - neighbouring H1', H2' and H2" protons and between the H8 base proton and the intraresidual H2' / H2" protons, while the distance between the H8 proton and the intraresidual H1' proton decreases. To a first approximation the cross peak intensities are inversely proportional to the sixth power of the distance between the respective proton spins. Hence a rotation of ~ $150^{o}$ about the glycosidic bond into a syn-type conformation provides a qualitative explanation of the effects observed in the NOESY spectra (Figs. 7.7 and 7.8). It is noted at this point that this conclusion can only be drawn for the bases at the end of the chain, i.e. 1, 2 and 6, which give rise to well resolved resonances. For the other bases this cannot (yet) be established unambiguously as a result of resonance overlap. In earlier studies from our laboratory [69] it was proposed that in the helical structure of the gene-5 protein - DNA complex the base planes are directed along the helical axis. Subsequent CD experiments by Scheerhagen et al. [147] suggested that such a structure is only possible when the bases are in a syn conformation. The present study indicates that at least part of the bases may occur in the syn conformation in the DNA gene-5 protein - complex.

There is no disagreement between our results and the two-dimensional $^{1}H$ NMR data of the gene-5 protein - oligonucleotide complexes, obtained by King

and Coleman [162], who conclude that the oligonucleotides possess an anti base conformation in the complex. At a $d(A)_4$ / protein monomer ratio of 1.5 (which corresponds to P / N = 0.166) King and Coleman do obtain internucleotide H8 - H1' NOEs while in a NOESY spectrum of a $d(A)_8$ - gene-5 protein mixture (ratio $d(A)_8$ / gene-5 protein = 0.5, so P / N = 0.25) only intranucleotide H8 - H1' NOEs are observed (Fig 7 in reference [162]). We attribute the absence of internucleotide H8 - H1' NOEs in this latter case as resulting from a syn base conformation So the picture emerges that the syn base conformation becomes apparent at least at large P / N ratios.

## 7.6 APPENDIX

In this appendix diagonal and cross peak intensities in two-dimensional exchange spectra are calculated using the theory developed by Jeener et al [152] as the starting point A model system is considered consisting of molecules of low and high molecular weight which can form complexes with each other and for which NOE cross peaks are observed It will be demonstrated that the cross peaks detected when exchange takes place between the two molecular forms are characteristic of the dipolar exchange in the high molecular weight form

### 7.6a Method to calculate peak intensities in two dimensional exchange spectra

In two-dimensional NMR exchange spectra the magnetisation can be transferred by direct dipolar interaction and by chemical exchange These effects can be made manifest in the spectrum by the following pulse sequence [152, 153,163] $90^0$ - evolution time ($t_1$) - $90^0$ - mixing time ($t_m$) - $90^0$ - detection time ($t_2$) In practice the $90^0$ pulses can be phased in such a manner as to enable quadrature detection At the beginning of the experiment the equilibrium magnetization (sum vector of the N spin system), $M_0$, points in the z-direction We define the complex transverse magnetization of nuclei of type $\ell$ by

$$M^+_\ell(t) = M_{\ell,x}(t) + i\, M_{\ell,y}(t) \qquad (7\ 1)$$

After the first $90^{o}$ pulse along the x-direction an N-dimensional transverse magnetization vector

$$M^{+}(0) = M_0 \tag{7.2}$$

is obtained. During the evolution period the precession, decay and exchange of the magnetization are described by the differential equation

$$(d/dt)\ M^{+} = i\ \Omega\ M^{+}(0) \tag{7.3}$$

in which the complex matrix $\Omega$ is the socalled spectral matrix (vide infra). Eq. 7.3 has the formal solution:

$$M^{+}(t_1) = \exp(i\Omega t_1)\ M^{+}(0) \tag{7.4}$$

A second $90^{o}$ pulse creates longitudinal magnetization:

$$M_z(t_1, t_m=0) = -\ \mathrm{Re}\{M^{+}(t_1)\} \tag{7.5}$$

During the mixing time (see pulse sequence), relaxation and exchange of magnetization is described by the following differential equation:

$$(d/dt)\ M_z = L\ \{M_z - M_0\} \tag{7.6}$$

in which L is the longitudinal relaxation matrix. The solution of this differential equation is (using Eqs. 7.3 - 7.6)

$$M_z(t_1, t_m) = [1 - \exp(L\ t_m)\ [\mathrm{Re}\{\exp(i\Omega t_1)\} + 1]]\ M_0 \tag{7.7}$$

The terms independent of $t_1$ give rise to the so-called axial peaks in the final spectrum and contain no information about the exchange process [149]. These peaks can be cancelled by a pulse phase alternation [164] and therefore the relevant part of Eq. 7.7 is

$$M_z(t_1, t_m) = -\ \exp(L\ t_m)\ \mathrm{Re}\{\ \exp(i\Omega t_1)\ \}\ M_0 \tag{7.8}$$

The third pulse converts $M_z(t_1,t_m)$ into transverse magnetization. As for the evolution period the time dependence of the magnetization during the detection period is governed by the spectral matrix. The final time dependence of the observable magnetization vector is given by

$$\Sigma_i[M^+(t_1,t_m,t_2)]_i =$$

$$-\Sigma_{i,j,\ell,p}\ [\ \exp(i\Omega t_2)]_{i,j}[\exp(Lt_m)]_{j,\ell}\quad \mathrm{Re}[\exp(i\Omega t_1)]_{\ell,p}\ [M_0]_p \qquad (7.9)$$

Exact numerical expressions for the multiplicands in Eq. 7.8 can only be obtained after diagonalization of the spectral and exchange matrices, i.e.

$$(i\Omega') = X^{-1}\ (i\Omega)\ X \text{ and } L' = Y^{-1}\ (L)\ Y \qquad (7.10)$$

where X and Y are the eigenvector matrices of $i\Omega$ and L respectively (see e.g. [165]). This leads to the following expressions for the matrix elements in Eq. 7.10:

$$[\exp(i\Omega t_2)]_{ij} = \Sigma_k\ [X]_{ik}\ [X^{-1}]_{kj}\ [\exp(i\Omega')\ t_2]_{kk} \qquad (7.11)$$

$$[\exp(i\Omega t_1)]_{\ell p} = \Sigma_n\ [X]_{\ell n}\ [X^{-1}]_{np}\ [\exp(i\Omega')\ t_1]_{nn} \qquad (7.12)$$

and

$$[\exp(L\ t_m)]_{j\ell} = \Sigma_m\ [Y]_{jm}\ [Y^{-1}]_{m\ell}\ [\exp(L'\ t_m)]_{mm} \qquad (7.13)$$

Two dimensional Fourier transformation then yields the complex spectrum

$$S^+(\omega_1,t_m,\omega_2) = -\Sigma_{i,j,k,\ell,m,n,p}\ [X]_{ik}\ [X^{-1}]_{kj}\ (i([\Omega']_{kk} - \omega_2))^{-1}\ [Y]_{jm}\ [Y^{-1}]_{m\ell}\ \exp(L't_m)_{mm}\ \mathrm{Re}\{[X]_{\ell n}\ [X^{-1}]_{np}\ (i([\Omega']_{nn} - \omega_1))^{-1}\ )[M_0]_p \qquad (7.14)$$

The 2D exchange spectrum consists of diagonal peaks (k=n) and cross peaks ($k\neq n$) at positions $(\omega_1,\omega_2) = (\mathrm{Re}\{[\Omega']_{nn}\}\ ,\ \mathrm{Re}\{[\Omega']_{kk}\}$ with the linewidts $\mathrm{Im}([\Omega']_{nn})$ and $\mathrm{Im}([\Omega']_{kk})$ and proportional to the mixing coefficient

$$a(n,k) = -\Sigma_{i,j,\ell,m,p} \quad [X]_{ik}[X^{-1}]_{kj}[Y]_{jm}[Y^{-1}]_{m\ell} \quad (\exp(L't_m)_{mm} \quad \mathrm{Re}\{[X]_{\ell n}[X^{-1}]_{np}\} \tag{7.15}$$

The eigenvectormatrix X changes to the unit matrix in the case of slow exchange because the off-diagonal elements of $i\Omega$ are zero in that case, see also Eq. 7.17. As a result $\Omega' = \Omega$ and

$$a(n,k) = -\Sigma_m \; [Y]_{km}[Y^{-1}]_{mn}(\exp(L't_m))_{mm} \qquad \text{(slow exchange)} \tag{7.16}$$

The peak-intensities in the 2D exchange spectrum of the model system defined below can now be calculated.

## 7.6b Model system

We consider the following model: a low molecular weight molecule carrying two spins may form a complex with a high molecular weight molecule (the protein) as indicated in the following reaction.

$$\text{protein} + \text{form A} \begin{matrix} 1 \\ \rightleftharpoons \\ 2 \end{matrix} \begin{matrix} k_{AB} \\ \rightleftharpoons \\ k_{BA} \end{matrix} \text{form B} \begin{matrix} 3 \\ \rightleftharpoons \\ 4 \end{matrix} \tag{7.17}$$

Form A is the low molecular weight molecule in the free state, form B represents this molecule complexed to the high molecular weight molecule. In the free state the spins are numbered 1 and 2, in the complexed state 3 (being 1 in the free state) and 4 (being 2 in the free state). The spectral and exchange matrices representative for this model system are given below.

$$i\Omega = \begin{pmatrix} -R_2(1)-\tau_{AB}^{-1}+i\,\omega(1) & 0 & k_{BA} & 0 \\ 0 & -R_2(2)-\tau_{AB}^{-1}+i\,\omega(2) & 0 & k_{BA} \\ k_{AB} & 0 & -R_2(3)-\tau_{BA}^{-1}+i\,\omega(3) & 0 \\ 0 & k_{AB} & 0 & -R_2(4)-\tau_{BA}^{-1}+i\,\omega(4) \end{pmatrix} \tag{7.18}$$

The precession frequency of nucleus $i$ is given by

$$\omega(i) = \omega_0 \, (1 + \delta(i) \, 10^{-6}) \qquad (7\ 19)$$

in which $\delta(i)$ is the chemical shift of nucleus $i$ (in the calculation we take $\delta(1) = 8\ 00$ ppm, $\delta(2) = 4\ 00$ ppm, $\delta(3) = 7\ 00$ ppm and $\delta(4) = 3\ 00$ ppm) and $\omega_0$ is the spectrometer frequency $2\pi\ 500\ 10^6$ rad $s^{-1}$) The transverse relaxation constant of nucleus $i$, $R_2(i)$, is given by [166]

$$R_2(i) = \qquad (7\ 20)$$
$$1\ 5\ K\ r^{-6}\ [3\tau_c(i) + 5\tau_c(i)\ /(1 + \omega_0{}^2\ \tau_c(i)^2) + 2\ \tau_c(i)/(1 + 4\omega_0{}^2\tau_c(i)^2)]$$

For protons the constant K has the value $5\ 6968$ $Å^6\ s^{-2}$ The exchange matrix equals

$$L = \begin{vmatrix} \rho(1)+\tau_{AB}{}^{-1} & \sigma(1) & -\tau_{BA}{}^{-1} & 0 \\ \sigma(2) & \rho(2)+\tau_{AB}{}^{-1} & 0 & -\tau_{BA}{}^{-1} \\ -\tau_{AB}{}^{-1} & 0 & \rho(3)+\tau_{BA}{}^{-1} & \sigma(3) \\ 0 & -\tau_{AB}{}^{-1} & \sigma(4) & \rho(4)+\tau_{BA}{}^{-1} \end{vmatrix} \qquad (7\ 21)$$

in which the spin-lattice relaxation is given by

$$\rho(i) = \Sigma_j \, (W_0(i,j) + 2W_1(i,j) + W_2(i,j)) \qquad (7\ 22)$$

and the cross-relaxation by

$$\sigma(i) = \Sigma_j \, (W_2(i,j) - W_0(i,j)) \qquad (i \neq j) \qquad (7\ 23)$$

The two spin system transition probabilities used in the former two Eqs are

$$W_0(i,j) = (K/r(i,j)^6)\ \tau_c(i)\ /\ [1 + \{\omega_0(\delta(i) - \delta(j))\tau_c(i)\}^2]$$
$$W_1(i,j) = (K/r(i,j)^6)\ 1\ 5\ \tau_c(i)\ /\ [1 + (\omega_0\ \tau_c(i))^2]$$
$$W_2(i,j) = (K/r(i,j)^6)\ 6\tau_c(i)\ /\ [1 + \{2\omega_0\ \tau_c(i)\}^2] \qquad (7\ 24)$$

In order to make the model calculations more realistic, an extra contribution of

$$K\,\tau_c\,\{(1/4)^6 + (1/3)^6\}\{1/[1 + (\omega_0\,\tau_c(1)^2] + 4/[1 + 4(\omega_0\,\tau_c(1))^2]\} \qquad (7\ 25)$$

is added to the spin-lattice relaxation (see Eqs 7 20 and 7 21), to indicate that the lattice relaxation only has a minor influence on the 2D exchange peak intensities The terms in Eq 7 25 correspond with the existence of two hypothetical spins at a distance of three and four Å (see [167] Now all input-parameters are given, and the 2D exchange peak intensities can be calculated by using Eq 7 15 and standard matrix manipulation routines [168]

For simplicity it has been assumed that the solution conditions are adjusted such that $\tau_{AB}^{-1} = \tau_{BA}^{-1} = k$ The dipolar relaxation only takes place between spin 1 and spin 2 in form A and between spin 3 and spin 4 in form B (besides some small 'leak' to the external spins, represented by Eq 7 25) The difference in molecular weight between the molecules is reflected in the rotational correlation times For form A it is assumed that $\tau_c = 2 \times 10^{-9}$ s and for form B that $\tau_c = 10^{-7}$ s The distance between the protons is the same in both forms and amounts 3 0 Å The results of the calculations are summarized in Fig 7 10 where the intensities of cross peaks and diagonal peaks have been plotted as a function of k In the absence of exchange the intensities of the cross peaks in a NOESY spectrum of form A are much smaller than the intensities of the cross peaks in a NOESY spectrum of form B a(1,2) = 0 03 (see curve c) and a(3,4) = 0 48 (see curve f) (Fig 7 10A) Due to the lower rotational correlation time dipolar exchange of magnetisation is less efficient in form A than it is in form B However, upon increasing the exchange rate the intensities shift to a common value of 0 20, while the resonances shift to a common position (shown in the upper part of Fig 7 10A) The peaks (1,3) and (2,4) (for both see curve d in Fig 7 10A) are the result of chemical exchange In the situation of absence of dipolar exchange in form A the intensities of diagonal and cross peaks are given in Fig 7 10B At higher exchange rates the intensity of cross peaks between protons in form A (a(1,2), curve c) increases to 0 19, despite the absence of direct dipolar exchange In this case the magnetisation is first chemically transferred from proton 1 to proton 3 and then chemically transferred from proton 3 to proton 4 due to dipolar exchange in form B This gives a cross peak at the same position and at the same intensity as when the magnetisation is subsequently transferred from proton 4 to proton 2 at high exchange rates The only reason for the appearance of cross peaks is the dipolar relaxation in form B The results of these mod-

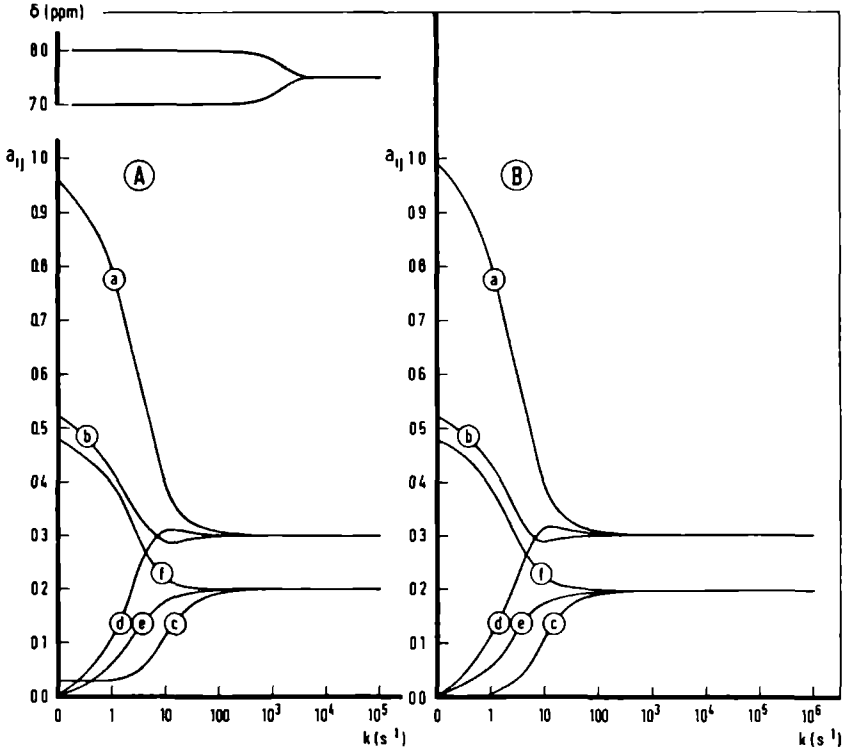



Fig 7 10 Simulation of the intensities $a_{ij}$ of the peaks in a 2D exchange spectrum of the four spin model system, given in the text, as a function of the chemical reaction rate k ($k_{AB} = k_{BA}$) in Eq 7 1 (Appendix) Two different cases are described A) A direct dipolar exchange exists between spins 1 and 2 and between spins 3 and 4 In the upper part of the figure the chemical shift positions of spins 1 and 3 are given B) A direct dipolar exchange exists only between spins 3 and 4 The letters a to f indicate the following combinations of the coupled spins i and j a (1,1), b (3,3), c (1,2), d (2,4), e (1,4) and (3,2), f (3,4) Note that $a_{ij} = a_{ji}$

el calculations clearly show that in the case of fast exchange between two molecular forms with a large difference in the effectiveness of dipolar relaxation the cross peaks in the exchange spectrum of a mixture of both forms originate from the molecular form with the greatest dipolar exchange. Furthermore, better spectra, that means in this case stronger cross peaks, are obtained for higher molecular weights, in contrast to most NMR experiments.

The results obtained here are analogous to those which Clore and Gronenborn derived for one dimensional exchange spectra [159]. On the basis of the results obtained in this Appendix it is concluded that the NOEs observed between the protons of $d(A)_6$ for the gene-5 protein - DNA sample arise from the nucleotide in the bound state and not from the free state.

## REFERENCES


[1] Hofschneider, P.H. and Preuss, A. (1963) J. Mol. Biol. 7, 450-451

[2] Webster, R.E., Grant, R.A. and Hamilton, L.A.W. (1981) J. Mol. Biol. 152, 357-374

[3] Van Wezenbeek, P.M.G.F., Hulsebos, T.J.M. and Schoenmakers, J.G.G.(1980) Gene 11, 129-148

[4] Wachtel, E.J., Wiseman, R.C., Pigram, W.J., Marvin, D.A. and Manualidis, L. (1974) J. Mol. Biol. 88, 601-618

[5] Banner, D.W., Nave, C. and Marvin, D.A. (1981) Nature 289, 814-816

[6] Marvin, D.A. and Wachtel, E.J. (1975) Nature 253, 306-315

[7] Marvin, D.A., Wiseman, R.L. and Wachtel, E.J. (1974) J. Mol. Biol. 82, 121-138

[8] Gray, C.W., Brown, R.S. and Marvin, D.A. (1981) 146, 621-627

[9] Goldsmith, M.E. and Koningsberg, W.H. (1977) Biochemistry 16, 2686-2694

[10] Smilowitz, H. (1974) J. Virol. 13, 94-99

[11] Kornberg, A. (1974) in: DNA synthesis (Freeman, W.H ed.) San Fransisco USA

[12] Geider, K. and Kornberg, A. (1974) J. of Biol. Chem. 249, 3999-4005

[13] Schaller, H., Uhlmann, A. and Geider, K. (1976) Proc. Nat. Acad. Sci. USA 73, 49-53

[14] Zinder, N.D. and Horiuchi, K , Microbiol. Rev. 49 (1985) 101-106

[15] Van Dorp, B., Schneck, P.K. and Staudenbauer, W.L. (1979) Eur. J. Biochem., 94, 445-450

[16] Ray, D.S., Bscheider, H-P. and Hofscheider, P.H. (1966) J. Mol. Biol. 21, 473-483

[17] Alberts, B. and Frey, L. (1970) Nature 227, 313-318

[18] Oey, J.L. and Knippers, R. (1972) J. Mol. Biol. 68, 125-138

[19] Mazur, B.J. and Model, P. (1973) J. Mol. Biol. 78, 285-300

[20] Pratt, D., Laws, P. and Griffith, J. (1974) J. Mol. Biol. 82, 425-439

[21] Ray, D.S. (1978) in: The Single-Stranded DNA Phages (Denhardt, D.T., Dressler, D. and Ray, D.S., eds.) Cold Spring Harbor Laboratory Monograph Series, pp. 325-339, Cold Spring Harbor, New York, USA

[22] Webster, R.E. and Cashman, J.S. (1978) in: The Single-Stranded DNA Phages (Denhardt, D.T., Dresiler, D. and Ray, D.S., eds.) Cold Spring Harbor Monograph Series, pp. 557-569, Cold Spring Harbor, New York, USA

[23] Peeters, B.P.H. (1985) thesis, Catholic University, Nijmegen, the Netherlands

[24] Ray, D.S. (1977) Comprehensive Virology volume 7, (Fraenkel-Conrat, H. and Wagner, R. eds. Plenum Press New York) 105-178

[25] Van den Hondel, C.A., Weijers, A., Konings, R.N.H. and Schoenmakers, J.G.G. (1975) Eur. J. Biochem. 53, 559-567

[26] Beck, E., Sommer, R., Auerswald, E.A., Kurz, Ch., Zink, B., Osterburg, G., Schaller, H., Sugimoto, K., Sugisaki, H., Okamoto, T. and Takanami, M. (1978) Nucleic Acids Res. 5, 4495-4503

[27] Hill, D.F. and Petersen, G.B. (1982) J. Virol. 44, 32-46

[28] Henry, T.J. and Pratt, D. (1969) Proc. Nat. Acad. Sci. USA 62, 800-807

[29] Day, L.A. (1973) Biochemistry 12, 5329-5339

[30] Cuypers, T., Van der Ouderaa, F.J. and De Jong, W W. (1974) Biochem. Biophys. Res. Commun. 59, 557-563

[31] Nakashima, Y., Dunker, A.K., Marvin, D.A. and Koningsberg, W. (1974) FEBS Letters 40, 290-292 (erratum (1974) FEBS Letters 43, 125)

[32] Rasched, I. and Pohl, F.M. (1974) Febs Letters 46, 115-118

[33] Pretorius, H.T., Klein, M. and Day, L.A. (1975) J. Biol. Chem. 250, 9262-9269

[34] Cavalieri, S.J., Neet, K. and Goldthwait, D.A. (1976) J. Mol. Biol. 102, 697-711

[35] Pörschke, D. and Rauh, H. (1983) Biochemistry 22, 4737-4745

[36] Cavalieri, S.J., Goldthwait, D.A. and Neet, K.E. (1976) J. Mol. Biol. 102, 713-722

[37] Alma, N.C.M., Harmsen, B.J.M., De Jong, E.A.M., Van de Ven, J. and Hilbers, C.W. (1983) J. Mol. Biol. 163, 47-62

[38] McPherson, A., Molineux, I. and Rich A. (1976) J. Mol. Biol. 106, 1077-1081

[39] McPherson, A., Jurnak, F., Wang, A., Kolpak, F., Molineux, I. and Rich, A. (1979) Cold Spring Harbor Symp. Quant. Biol 43, 21-28

[40] McPherson, A., Jurnak, F., Wang, A., Kolpak, F., Rich, A., Molineux, I. and Fizgerald, P. (1980) Bioph. J., 32, 155
[41] Brayer, G.D. and McPherson, A. (1983) J. Mol. Biol. 169, 565-596
[42] Pratt, D., Erdahl, W.S., J. Mol. Biol. 37 (1968) 181-200
[43] Tyaglov, B.V., Naidenova, N.M., Minaev, V.E. and Permogorov, V.I. (1980) Mol. Biol. (Russian) 14, 814-819
[44] Hollwitt, E. and Krasna, A.I. (1982) Arch. Bioch. Bioph. 214, 792-805
[45] Veiko, N.N., Gromova, E.S. and Shabarova, Z.A. (1978) Mol. Biol. (Russian) 13, 1136-1146
[46] Geider, K. (1978) Eur. J. Biochem. 87, 617-622
[47] Model, P., Mcgill, C., Mazur, B. and Fulford, W.D. (1982) Cell 29, 329-335
[48] Yen, T.S.B. and Webster, R.E. (1982) Cell 29, 337-345
[49] Fulford, W. and Model, P. (1984) J. Mol. Biol. 178, 137-153
[50] Torbet, J., Gray, D.M., Gray, C.W., Marvin, D.A. and Siegrist, H. (1981) J. Mol. Biol. 146, 305-320
[51] Gray, C.W., Kneale, G.G., Leonard, K.R., Siegrist, H. and Marvin, D.A. (1982) Virology 116, 40-52
[52] Gray, D.M., Gray, C.W. and Carlson, R.D. (1982) Biochemistry 21, 2702-2713
[53] Berkowitz, S.A. and Day, L.A. (1976) J. Mol. Biol. 102, 531-547
[54] Marvin, D.A. (1978) in: Single-Stranded DNA phages (Denhardt, D.T., Ray, D.S. and Dressler, D., eds.) Cold Spring Harbor Laboratory Monograph Series, pp. 583-603, Cold Spring Harbor, New York, USA
[55] Day, L.A. and Wiseman, R.L. (1978) in: The single-stranded DNA phages, pp. 605-626, Cold Spring Harbor Lab., Cold Spring Harbor, N.Y.
[56] Saenger, W. (1984) in: Principles of Nucleic Acid Structure, Springer Advanced Text in Chemistry, Springer-Verlag, New York, USA
[57] McPherson, A., Wang, A.H.J., Jurnak, F.A., Molineux, I., Kolpak, F. and Rich, A. (1980) J. Biol. Chem. 255, 3174-3177
[58] Brayer and G.D., McPherson A. (1984) Biochemistry 23, 340-349
[59] Shimamoto, N. and Utiyama, H. (1983) Biochemistry 22, 5869-5878
[60] Garssen, G.J., Hilbers, C.W., Schoenmakers, J.G. and Van Boom, J.H. (1977) Eur. J. Biochem. 81, 453-463

[61] Anderson, R.A., Nakashima, Y. and Coleman, J.E. (1975) Biochemistry, 14, 907-916
[62] Coleman, J.E., Anderson, R.A., Ratcliffe, R. and Armitage, I.A. (1976) Biochemistry 15, 5419-5430
[63] Bobst, A.M., Ireland, J.C. and Bobst, E.V. (1984) J. Biol. Chem. 259, 2130-2134
[64] Paradiso, P.R. and Konigsberg, W. (1982) J. Biol. Chem. 257, 1462-1467
[65] Brayer, G.D. and McPherson, A. (1984) J. Biomol. Struct. Dyn. 2, 495-510
[66] Brayer, G.D. and McPherson, A. (1985) Eur. J. Biochem. 150, 287-296
[67] Alma, N.C.M., Harmsen, B.J.M., Van Boom, J.H., Van der Marel, G. and Hilbers, C.W. (1982) Eur. J. Biochem. 122, 319-326
[68] Gray, D., Sherry, A.D., Teherani, J. and Kansy, J.W. (1984) J. Biomol. Struct. Dyn. 2, 77-91
[69] Alma, N.C.M., Harmsen, B.J.M., Van Boom, J.H., Van der Marel, G. and Hilbers, C.W. (1983) Biochemistry 22, 2104-2115
[70] Kansy, J.W., Clack, B.A. and Gray, D.M. (1986) J. Biomol. Struct. Dyn.3, 1079-1110
[71] McGhee, J.D. and Von Hippel, P.H. (1974) J. Mol. Biol. 86, 469-489
[72] Dunker, A.K. and Anderson, E.A. (1975) Biochim. Biophys. Acta, 401, 31-34
[73] Record, M.T., Anderson C.F. and Lohman, T.M. (1978) Q. Rev. Biophys. 11, 103-178
[74] Record, M.T., jr., Lohman, T.M. and De Haseth, P. (1976) J. Mol. Biol. 107, 145-158
[75] Alma, N.C.M., Harmsen, B.J.M., Hilbers, C.W., Van der Marel, G. and Van Boom, J.H. (1981) FEBS Letters 135, 15-20
[76] Alma, N.C.M., Harmsen, B.J.M., Hull, W.E., Van der Marel, G. and Van Boom, J.H., Hilbers, C.W. (1981) Biochemistry 20, 4419-4428
[77] Coleman, J.E. and Armitage, E.M. (1978) Biochemistry 17, 5038-5045
[78] Garssen, G.J., Tesser, G.I., Schoenmakers, J.G.G. and Hilbers, C.W. (1980) Biochim. Biophys. Acta 607, 361-371
[79] O'Connor, T.P. and Coleman, J.E. (1983) Biochemistry 22, 3375-3381
[80] Hilbers, C.W., Garssen, G.J., Kaptein, R., Schoenmakers, J.G.G., and Van Boom, J.H. (1978) in: Nuclear Magnetic Resonance

Spectroscopy in Molecular Biology, pp. 351-364, B. Pullman (ed), Reidel Publishing Company, Dortrecht, Holland

[81] Garssen, G.J., Kaptein, R., Schoenmakers, J.G. and Hilbers, C.W. (1978) Proc. Natl. Acad. Sci. USA 75, 5281-5285

[82] Paradiso, P.R., Nakashima, Y. and Koningsberg, W. (1979) J. Biol. Chem. 254, 4739-4744

[83] Fulford, W. and Model, P. (1984) J. Mol. Biol 173, 211-226

[84] Kneale, G.G. and Wijnaendts van Resandt, R.W. (1985) Eur. J. Biochem. 149, 85-93

[85] Schwartz, G., Klose, S. and Balthasar, W. (1970) Eur. J. Biochem. 12, 442-467

[86] Schwartz, G. (1977) Biophys. Chem. 6, 65-76

[87] Epstein, I.R. (1978) Biophys. Chem. 8, 327-339

[88] Schnarr, M., Durand, M. and Maurizot, J.-C. (1983) Biochemistry 22, 3563-3570

[89] Schwartz, G. and Stankowski, S. (1979) Biophys. Chem. 10, 173-181

[90] Schwartz, G. and Watanabe, F. (1983) J. Mol. Biol. 163, 467-484

[91] Kowalczykowski, S.C., Lonberg, N., Newport, J.W. and Von Hippel, P.H. (1981) J. Mol. Biol. 145, 75-104

[92] Lohman, T.M. and Kowalczykowski, S.C. (1981) J. Mol. Biol. 152, 67-109

[93] Epstein, I.R. (1979) Biopolymers 18, 2037-2050

[94] Lohman, T.M. (1984) Biochemistry 23, 4656-4665

[95] Balazs, A.C. and Epstein, I.R. (1984) Biopolymers 23, 1249-1259

[96] Lohman, T.M. (1984) Biochemistry 23, 4665-4675

[97] Steitz, T.A., Weber, I.T., Ollis, D. and Brick, P. (1983) J. Biomol. Struct. Dyn. 1, 1023-1037

[98] Pabo, C.O. and Lewis, M. (1982) Nature 298, 443-447

[99] Paci, M., Pon, C.L., Losso, M.A. and Gualerzi, C.O. (1984) Eur. J. Biochem. 138, 193-200

[100] Tanaka, I., Appelt, K., Dijk, J., White, S.W. and Wilson, K.S. (1984) Nature 310, 376-381

[101] Lohman, T.M. (1983) Biopolymers 22, 1697-1713

[102] Newport, J.W., Lonberg, N., Kowalczykowski, S.C. and Von Hippel, P.H. (1981) J. Mol. Biol. 145, 105-121

[103] Watanabe, F. and Schwarz, G. (1983) J. Mol. Biol. 163, 485-498

[104] Gray, C.W., Page, G. and Gray, D.M. (1984) J. Mol. Biol. 175, 553-559
[105] Von Hippel, P.H., Kowalczykowski, S.C., Lonberg, N., Newport, J.W., Paul, L.S., Stormo, G.D. and Gold, L. (1982) J. Mol. Biol. 162, 795-818
[106] Konings, R.N.H., Jansen, J., Cuypers, T. and Schoenmakers, J.G.G.(1973) J. Virology 12, 1466-1472
[107] Holley, W. (1963) Biochem. Biophys. Res. Comm. 10, 186-188
[108] Gillam, I., Millward, S., Blew, D., Von Tigerstrom, M., Wimmer, E. and Tener, G.M. (1967) Biochemistry 6, 3043-3056
[109] Nishimura, S. in: Procedures in Nucleic Acid Research (Cantoni, G.L. and Davis, D.R. eds.), Harper and Row, New York, Vol. 2, 542-564
[110] Holmes, W.M., Hurd, R E., Reid, B.R., Ruemerman, R.A. and Hatfield, G.W. (1975) Proc. Natl. Acad. Sci. USA 72, 1068-1071
[111] Salemink, P.J.M., Raué, H.A., Heerschap, A., Planta, R.J. and Hilbers, C.W. (1981) Biochemistry 20, 265-272
[112] Miles, H.T. and Frazier, J. (1978) J. Am. Chem. Soc. 100, 8037-8038
[113] Howard, F.B. and Miles, H T (1982) Biochemistry 21, 6736-6745
[114] Hélène, C. and Lancelot, G. (1982) Prog. Biophys. Molec. Biol. 39, 1-68
[115] Record, M.T., Mazur, S.J., Melançon, P., Roe, J.-H., Shaner, S.L. and Unger, L. (1981) Ann. Rev. Biochem. 50, 997-1024
[116] Lohman, T.M., De Haseth, P.L. and Record, T.M., jr. (1980) Biochemistry 19, 3522-3530
[117] Walters, J.A.L.I., Geerdes, H.A.M. and Hilbers, C.W. (1977) Biophys. Chem. 7, 147-151
[118] Leroy, J.L. and Gueron, M. (1977) Biopolymers 16, 2429-2446
[119] Krauss, G., Sinderman, H., Schomburg, U. and Maass, G. (1981) Biochemistry 20, 5346-5352
[120] Pörschke, D. and Eggers, F. (1972) Eur. J. Biochem. 26, 490-498
[121] Edens, L., Konings, R.N.H. and Schoenmakers, J.G.G. (1978) Virology 86, 354-367
[122] Edens, L., Konings, R.N.H. and Schoenmakers, J.G.G. (1978) J. Virology 28, 835-842
[123] Shen, C-K., J., and Hearst, J.E. (1976) Proc. Natl. Acad. Sci. USA 73, 2649-2653

[124] Gray, C.P., Sommer, R., Polke, C., Beck, E. and Schaller H. (1978) Proc. Natl. Acad. Sci. USA 75, 50-53

[125] Schaller, H. (1979) Cold Spring Harbor Symposia on Quantitative Biology, XLIII-1, 401-408

[126] Huang, C-C. and Hearst, J.E. (1981) Nucleic Acids Research 9, 5587-5599

[127] Smits, M.A. (1982) thesis, Katholieke Universiteit Nijmegen, The Netherlands

[128] Tibanyenda, N., De Bruin, S.H., Haasnoot, C.A.G., Van der Marel, G.A., Van Boom, J.H. and Hilbers, C.W. (1984) Eur. J. Biochem. 139, 19-27

[129] Borer, P.N., Dengler, B., Tinoco, I. Jr. and Uhlenbeck, O.C., J. (1974) Mol. Biol. 86, 843-853

[130] Hilbers, C.W., Haasnoot, C.A.G., De Bruin, S.H., Joordens, J.J.M., Van der Marel, G.A. and Van Boom, J.H. (1985) Biochimie 67, 685-695

[131] Shen, C.J., Ikoku, A. and Hearst, J.E. (1979) J. Mol. Biol. 127, 163-175

[132] Doherty, D.H., Gauss, P. and Yord, L. (1982) in: Multifunctional proteins (Kane, J.F. ed.) C.R.C. Press Cleveland

[133] Russel, M.L., Gold. L., Morrissett, H. and O'Farell, P.Z. (1976) J. Biol. Chem. 251, 7263-7270

[134] Alberts, B., Frey, L. and Delius, H. (1972) J. Mol. Biol. 68, 139-152

[135] Greulich, K.O., Wijnaendts van Resandt, R W. and Kneale, G.G. (1985) Eur. J. Biophys. 11, 195-201

[136] Bobst, A.M., Langemeier, P.W., Warwick-Koochaki, P.E., Bobst, E.V and Ireland, J.C. (1982) J. Biol. Chem. 257, 6184-6193

[137] Cassani, G.R. and Bollum, F.J. (1969) Biochemistry 8, 3928-3936

[138] Wijnaendts van Resandt, R.W., Vogel, R.H. and Provencher, S.W. (1982) Rev. Sci. Instrum. 53, 1392-1397

[139] Yguerabide, J. (1972) Methods Enzymol. 26, 498-578

[140] Wahl, P. (1975) in: Biochemical Fluorescence Concepts (Chen, R.F. and Edelhoch, H. eds) vol. 1, pp. 1-44, Marcel Dekker, New York, USA

[141] Cantor, C.R. and Schimmel, P.R. (1980) in: Biophysical Chemistry, part 2, Techniques for the study of biological structure and function, pp. 433-465, W.H. Freeman and company, San Fransisco USA

[142] Vogel, R., SPLMOD, EMBL Data Processing Dept., EMBL Heidelberg, Western Germany (available on request)

[143] Kansy, J.W., Clack, B.A. and Gray, D.M. (1985, june) Fourth Conversation in Biomolecular Sterodynamics, Albany, New York, abstracts, pp. 217-218

[144] Gray, C.W. (1985, june) Fourth Conversation in Biomolecular Sterodynamics, Albany, New York, abstracts, pp. 219-220

[145] Römer, R., Schomburg, U., Krauss, G. and Maas, G. (1984) Biochemistry 23, 6132-6137

[146] Forsheit, A.B., Ray, D.S. and Lica, L. (1971) J. Mol. Biol. 57, 117-127

[147] Scheerhagen, M.A., Blok, Joh. and Van Grondelle, R. (1985) J. Biomol. Struct. Dyn. 2, 821-829

[148] Arentzen, R., Van Boeckel, C.A.A., Van der Marel, G. and Van Boom, J.H. (1979) Synthesis 137-139

[149] Aue, W.P., Bartholdi, E. and Ernst, R.R. (1976) J. Chem. Phys. 64, 2229-2246

[150] Nagayama, K., Wüthrich, K. and Ernst, R.R. (1979) Biochem. Biophys. Res. Commun. 90, 305-311

[151] Bax, A. and Freeman, R. (1981) J. Magn. Reson. 44, 542

[152] Jeener, J., Meier, B.H., Bachmann, P. and Ernst, R.R. (1979) J. Chem. Phys. 71, 4546-4553

[153] Kumar, A., Wagner, G., Ernst, R.R. and Wüthrich, K. (1980) Biochem. Biophys. Res. Commun. 96, 1156-1163

[154] Haasnoot, C.A.G., Van de Ven, F.J.M. and Hilbers, C.W. (1984) J. Magn. Res. 56, 343-349

[155] Olsthoorn, C.S.M., Bostelaar, L.J., Van Boom, J.H. and Altona, C. (1980) Eur. J. Biochem. 112, 95-110

[156] Haasnoot, C.A.G., Westerink, H.P., Van der Marel, G.A. and Van Boom, J.H. (1983) J. Biomol. Struct. Dyn. 1, 131-149

[157] Hare, D.R., Wemmer, D.E., Chou, S.H., Drobny, G. and Reed, B. (1983) J. Mol. Biol. 171, 319-336

[158] Scheek, R.M., Boelens, R., Russo, N., Van Boom, J.H. and Kaptein, R. (1984) Biochemistry 23, 1371-1376

[159] Clore, G.M. and Gronenborn, A.M. (1982) J. Magn. Reson. 48, 402-417

[160] Olsthoorn, C.S.M. (1981), thesis University Leiden, the Netherlands

[161] Arnott, S., Campbell Smith, P.J. and Chandrasekaran, R. (1976) Handbook of Biochemistry and Molecular Biology, CRC Cleveland
[162] King, G.C. and Coleman, J.E. (1987) Biochemistry 26, 2929-2937
[163] Kumar, A., Ernst, R.R. and Wüthrich, K. (1980) Biochem. Biophys. Res. Commun. 95, 1-6
[164] Macura, S. and Ernst, R.R. (1980) Mol. Phys. 41, 95-117
[165] Perrin, C.L. (1970) Mathematics for Chemists (Wiley-Interscience) New York, USA
[166] Solomon, I. (1955) Physical Review 99, 559-565
[167] Kalk, A. and Berendsen, H.J.C. (1976) J. Magn. Reson. 24, 343-366
[168] NAg (1984), Fortran 77 Nag library routines

## SUMMARY

The interaction of proteins with nucleic acids plays a central role in cellular processes. In this thesis the interaction of the bacteriophage M13 encoded gene-5 protein with nucleic acids has been chosen as a model system to study the principles underlying protein - nucleic acid interactions. Much is already known of the life cycle of bacteriophage M13 and the role of its gene-5 protein in this process. Chapter 1 gives a literature review of the knowledge collected so far.

It appears that the binding of gene-5 protein to DNA or RNA can be described by the model of McGhee and Von Hippel, in which the nucleic acid is considered as an infinite lattice of repeating units to which a protein molecule can bind covering a sequence of three or four of these units. Binding theories needed to derive binding constants, stoichiometry and cooperativity of the binding are given in Chapter 2, including extension or modification of the McGhee and Von Hippel model for the binding to smaller lattices (oligonucleotide binding) or binding in more than one specific mode. This chapter also describes some theoretical aspects of the kinetics of the binding, especially concerning the irreversible dissociation of complexes of the protein with poly- or oligonucleotides.

The influence of the type of sugar or base in the DNA or RNA on the binding properties was studied in Chapter 3 by means of fluorescence titrations with salt of complexes of gene-5 protein with various homo- and heteropoly(deoxy)ribonucleotides. It appears that the base and sugar composition of the polynucleotides are the major factors which determine the binding specificity. In contrast the salt dependency of the binding and the cooperativity are equal within the experimental accuracy for all the polynucleotides studied. The significance of these binding characteristics for the role of the protein in vivo is discussed. From calculations based upon these binding data it follows that a functional hairpin (hairpin A) in the intergenic region of the M13 genome is not melted by gene-5 protein, in line with suggestions in the literature. On the other hand, in the experiments of Fulford and Model [83] it has been shown that the gene-5 protein and the gene-32 protein (encoded by bacteriophage T4) cannot replace each other translational repressor functions.

Calculations based on the binding data known for gene-32 protein and on those given for gene-5 protein in Chapter 3 reveal that an explanation for the results of Fulford and Model cannot be given by considering the binding affinities alone, but that some preferential binding to a certain structure (or sequence) of mRNA will also play a role.

Chapter 4 describes fluorescence depolarisation studies of the binding of gene-5 protein to oligo(deoxy)thymidilic acids. From the dependence of the rotational correlation time, obtained from these experiments on the nucleotide to protein ratio it was evident that the cooperativity of the binding in these complexes is very low or even completely absent, which is in sharp contrast to the polynucleotide binding. Also the stoichiometry of the binding appeared to be lower for oligonucleotides than for polynucleotides (3 residues per protein molecule instead of 4).

In Chapter 5 stopped flow experiments are presented which were used to study the dissociation kinetics of gene-5 protein - poly(dA) and M13 DNA complexes upon addition of salt. For the complexes with poly(dA) always zeroth - order kinetics is obtained independent of the initial degree of saturation of the lattice, while for the M13 DNA complex this is only valid at complete saturation of the complex before dissociation. At lower values of the initial degree of saturation first - order kinetics is observed. This behaviour can be explained only partly by the existing kinetic models for this type of dissociation, which are described in Chapter 2. However, a complete explanation can be obtained by assuming that only protein molecules can dissociate which are bound on isolated sites on the lattice. This dissociation is accompanied by a redistribution of bound protein, which is fast with respect to the dissociation itself for the poly(dA) complex, but slow for the M13 DNA complex. NMR experiments of gene-5 protein - poly(dT) complexes showed that due to the high binding cooperativity this redistribution does not lead to line broadening of nucleotide resonances in the NMR spectra of these complexes.

Chapter 6 starts with fluorescence titrations of poly(dA) or M13 DNA with gene-5 protein as a function of the ionic strength, including very low values of the ionic strength. It appeared that the binding of the protein to these polynucleotides takes place in two different modes, of which one is only im-

portant at low salt concentrations. In addition this chapter contains an extensive study of the binding to oligonucleotides by means of fluorescence titrations, fluorescence depolarisation experiments and measurements of the irreversible dissociation kinetics of oligo(dT) - gene-5 protein complexes. The observed first order dissociation kinetics is consistent with the absence of binding cooperativity which already was concluded from the fluorescence depolarisation experiments. Furthermore comparison of the calculated and measured fluorescence depolarisation curves and the observation of a nucleotide chain length dependence of the apparent binding constant and of the dissociation rate constant all indicate that the protein is bound in these complexes with its two dimer halves to the same strand in stead of to two different strands as is generally accepted

Finally in Chapter 7 a two dimensional NMR study is presented of the gene-5 protein - $d(A)_6$ complex. A fast exchange is observed between the protein and the hexa(deoxy)adenylic acid on the NMR time scale. It is shown that under the experimental circumstances the cross peaks in the NOESY spectra of the hexa(deoxy)nucleotide arise from the $d(A)_6$ in the complexed state. Evidence is obtained that the nucleotide undergoes a conformational change upon binding of the protein.

De interactie tussen eiwitten en nucleinezuren speelt een centrale rol in cellulaire processen. In dit proefschrift is de interactie tussen het door bacteriofaag M13 gecodeerde gen-5 eiwit en nucleinezuren gekozen als model systeem om de grondbeginselen van de eiwit - nucleinezuur interacties te bestuderen voor enkelstrengs DNA bindende eiwitten. Veel is al bekend van de levenscyclus van bacteriofaag M13 en de rol van het eigen gen-5 eiwit in dit proces. Hoofdstuk 1 geeft een literatuuroverzicht van de tot op heden verzamelde kennis.

Het blijkt dat de binding van gen-5 eiwit aan DNA of RNA beschreven kan worden met het model van McGhee en Von Hippel. Hierin wordt het nucleinezuur beschouwd als een oneindig lang rooster van repeterende eenheden waaraan een eiwitmolecuul kan binden, daarbij drie of vier van deze eenheden bedekkend. In Hoofdstuk 2 worden bindingstheorieen gegeven die nodig zijn om bindingsconstanten en stoichiometry en coöperativiteit van de binding af te leiden. Ook wordt het McGhee en Von Hippel model aangepast c.q. uitgebreid voor de binding aan kleine roosters (oligonucleotide binding) en voor de situatie dat gelijktijdig binding op verschillende manieren plaatsvindt. Dit hoofdstuk beschrijft ook enige theoretische aspecten van de kinetiek van de binding, in het bijzonder aangaande de irreversibele dissociatie van complexen van het eiwit met poly- of oligonucleotiden.

De invloed van het type suiker of base in het DNA of RNA op de bindingseigenschappen wordt beschreven in Hoofdstuk 3 aan de hand van fluorescentie titraties met zout van complexen van gen-5 eiwit met verschillende homo- en heteropoly(deoxy)ribonucleotiden. Het blijkt dat de base- en suiker-samenstelling van de polynucleotiden de belangrijkste factoren zijn die de bindingsspecificiteit bepalen. Daarentegen blijken de zoutafhankelijkheid van de binding en de coöperativiteit hetzelfde te zijn binnen de experimentele meetnauwkeurigheid voor alle onderzochte polynucleotiden. De betekenis van deze bindingseigenschappen met betrekking tot de biologische rol van het eiwit wordt besproken. Uit berekeningen gebaseerd op deze bindingsgegevens volgt dat een functionele hairpin (hairpin A) in het intergeen gebied van het M13 genoom niet door

gen-5 eiwit wordt uitgesmolten, overeenkomstig suggesties in de literatuur. Aan de andere kant laten de experimenten van Fulford en Model [83] zien dat het gen-5 eiwit en het gen-32 eiwit (gecodeerd door bacteriofaag T4) elkaars translatie repressor functie niet kunnen overnemen. Berekeningen, gebaseerd op bindingsgegevens gepubliceerd voor gen-32 eiwit en eveneens bepaald voor gen-5 eiwit (zie Hoofdstuk 3) laten zien dat de resultaten van Fulford en Model niet verklaard kunnen worden door alleen maar de bindingsaffiniteiten te beschouwen, maar dat voorkeursbinding aan een bepaalde structuur (of sequentie) aan het mRNA ook een rol moet spelen.

Hoofdstuk 4 beschrijft fluorescentie depolarisatie studies betreffende de binding van gen-5 eiwit aan oligo(deoxy)thymidine zuren. Uit het verband tussen de rotatie-correlatietijd en de nucleotide - eiwit verhouding werd het duidelijk dat de coöperativiteit van de binding in deze complexen erg laag of zelfs in het geheel afwezig is, hetgeen in scherpe tegenstelling staat tot de polynucleotide binding. Ook de stoichiometry van de binding bleek lager voor oligonucleotiden dan voor polynucleotiden (3 residueen per eiwitmolecuul in plaats van 4).

In Hoofdstuk 5 worden stopped flow experimenten beschreven aangaande de dissociatie kinetiek van gen-5 eiwit - poly(dA) en M13 DNA complexen bij toevoeging van zout. Bij de complexen met poly(dA) wordt altijd een nulde orde kinetiek verkregen, onafhankelijk van de beginwaarde van de verzadigingsgraad van het rooster, terwijl voor het M13 DNA complex dit alleen zo is bij complete verzadiging van het complex voor dissociatie. Bij lagere beginwaarden wordt een eerste orde kinetiek waargenomen. Dit gedrag kan slechts gedeeltelijk verklaard worden met de bestaande kinetische modellen voor dit soort dissociatie zoals beschreven in Hoofdstuk 2. Een complete verklaring kan echter verkregen worden door te veronderstellen dat alleen eiwitmoleculen kunnen dissocieeren die gebonden zijn op geisoleerde plaatsen op het rooster. Samen met deze dissociatie vindt een herverdeling van gebonden eiwitten plaats, die voor het poly(dA) complex snel is in verhouding tot de dissociatie zelf, maar voor het M13 DNA complex langzaam. NMR experimenten aan gen-5 eiwit - poly(dT) complexen laten zien dat als gevolg van de hoge bindingscooperativiteit de herverdeling niet tot lijnverbreding van de nucleotide resonances in het NMR spectrum van deze complexen leidt.

Hoofdstuk 6 begint met fluorescentie titraties van poly(dA) en M13 DNA met gen-5 eiwit als functie van de zoutsterkte, ook voor zeer lage waarden van de zoutsterkte. De binding van het eiwit aan polynucleotiden blijkt plaats te vinden op twee verschillende manieren, waarvan er één alleen belangrijk is bij zeer lage zoutsterkte. Verder bevat dit hoofdstuk een uitgebreide studie van de binding aan oligonucleotiden door middel van fluorescentie titraties, fluorescentie depolarisatie experimenten en bepaling van de irreversibele dissociatie kinetiek van oligo(dT) - gen-5 eiwit complexen. Het feit dat eerste orde dissociatie kinetiek wordt gevonden komt overeen met de afwezigheid van bindingscoöperativiteit, hetgeen al geconcludeerd werd uit fluorescentie depolarisatie metingen. Vervolgens wijzen de onderlinge vergelijking van de berekende en gemeten fluorescentie depolarisatie curven alsmede het waarnemen van een nucleotide ketenlengte afhankelijkheid van zowel de dissociatiesnelheidsconstante als de bindingsconstante steeds op een bindingswijze van het eiwit waarbij de twee dimeerhelften met dezelfde nucleotide streng verbonden zijn, in plaats van met twee verschillende strengen, zoals algemeen wordt aangenomen.

Tenslotte wordt in Hoofdstuk 7 een twee-dimensionale NMR studie gepresenteerd van het gen-5 eiwit - $d(A)_6$ complex. Een snelle uitwisseling op de NMR tijdsschaal wordt waargenomen tussen het eiwit en het hexa(deoxy)adenine zuur. Onder de experimentele omstandigheden zijn de kruispieken in de NOESY spectra van het hexa(deoxy)nucleotide afkomstig van het $d(A)_6$ in de gebonden toestand. Bewezen wordt dat het nucleotide een conformatie-verandering ondergaat bij binding van het eiwit.

# DANKWOORD

Allereerst wil ik mijn ouders danken voor de mogelijkheden die zij mij geboden hebben om een studie te volgen. Daarnaast wil ik diegenen dank zeggen die in artikel II.c.1 van het promotie-regelement van de Katholieke Universiteit in Nijmegen bedoeld worden, voor hun begeleiding en stimulering. Verder dank ik Carla Prinse voor haar practische ondersteuning en introductie in het kweken van de bacteriofagen en de daarop volgende isolatie van het gen-5 eiwit. Ben Zweers, Odille Steijger, Harrie Marsman, Wim Engels, Geert-Jan Roelofs, Wim Schielen, Jan-Huib Franssen, Jacques Versteijnen, Bertus Welzen en Paul de Wild dank ik voor de bijdragen die zij geleverd hebben in het kader van hun stage, met name wat betreft de fluorescentie- en de stopped-flow experimenten. De kamergenoten Evert de Jong en Frank van de Ven hebben mij wegwijs gemaakt in de biofysische chemie en fungeerden vaak als klankbord. Verder wil ik danken Arend Heerschap, Henny Jansen, Jan-Rempt Mellema, Marcel Blommers, Theo Bosman, Marcel Akkermans en alle overige collega's binnen de afdelingen biofysische chemie, moleculaire biologie en moleculaire spectroscopie die mij met raad en daad ter zijde hebben gestaan. Zonder de technische ondersteuning van Cees Haasnoot, Pieter van Daal, Jos Joordens en Jan van Osch zou uitvoering van de NMR experimenten niet mogelijk geweest zijn. Roel Wijnaendts van Resandt dank ik voor de nuttige en gezellige twee weken die ik op het EMBL in Heidelberg heb doorgebracht. De hand en span diensten van John Roef waardeer ik minstens zo sterk als zijn goede humeur. Tenslotte dank ik de afdelingen Illustratie en Fotografie voor het verzorgen van de figuren en Theo Bosman voor zijn adviezen bij de tekstverwerking.

henk

# CURRICULUM VITAE

Henk Bulsink werd op 16 januari 1956 geboren te Dinxperlo. Hij behaalde in 1972 het MAVO diploma aan de Chr. Mavo Prinses Irene te Dinxperlo en in 1975 het VWO diploma aan de Chr. Scholengemeenschap te Aalten. Hierna werd begonnen met de studie scheikunde aan de Vrije Universiteit te Amsterdam. Het kandidaatsexamen werd in augustus 1978 afgelegd. Het doctoraalexamen scheikunde met als hoofdvak Fysische Chemie (prof. dr. ir. C. MacLean, onderwerp: kernspinresonantie) en de bijvakken Analytische Chemie (prof. dr. R.W. Frei en prof. dr. U.A.Th. Brinkman, onderwerp: vloeistofchromatografie) en Informatica (prof. dr. R.P. van de Riet) werd in april 1982 afgelegd. Tijdens de scheikunde studie vervulde hij assistentschappen op het eerste jaars scheikunde practicum en op het natuurkunde werkcollege voor biologen. De opleiding tot eerste-graads leraar scheikunde werd tevens gevolgd en gecombineerd met een invalperiode als leraar scheikunde aan de Chr. Scholengemeenschap Amsterdam-Noord. Van 1 mei 1982 tot 1 mei 1986 was hij, in dienst van de Nederlandse Organisatie voor Zuiver Wetenschappelijk Onderzoek (ZWO), werkzaam als promovendus op de afdeling Biofysische Chemie van de Katholieke Universiteit te Nijmegen. Gedurende deze periode werd het in dit proefschrift beschreven onderzoek verricht onder leiding van prof. dr. C.W. Hilbers en dr. B.J.M. Harmsen. Vanaf 1 september 1986 is hij werkzaam als ontwikkelaar in de werkgroep Materiaalkunde van het Centraal Laboratorium Glas van Philips in Eindhoven.

# STELLINGEN

behorende bij het proefschrift "The nucleic acid binding properties of the gene-5 protein encoded by bacteriophage M13".

1. Zowel Kneale en Wijnaendts van Resandt als Tsugita en Kneale hebben bindingsexperimenten verricht met het PF1 DNA bindende eiwit en verscheidene polynucleotiden. De indruk wordt gewekt dat deze binding coöperatief verloopt. Hiervoor worden echter te weinig indicaties gegeven. Duidelijkheid hieromtrent zou wel verkregen kunnen worden uit bindingscurven gemeten bij hogere zoutsterkte of uit Scatchard plots.

   Kneale, G.G. en Wijnaendts van Resandt, R.W. (1985) Eur. J. Biochem. 149, 85-93
   Tsugita, A. en Kneale, G.G. (1985) Biochem. J. 228, 193-199

2. Bij bindingsexperimenten van enkelstrengs nucleïnezuur bindende eiwitten met poly(dI) of poly(rI) dient rekening gehouden te worden met de helix-vorming van deze nucleotiden onder invloed van alkalimetaalionen.

   Newport, J.W., Lonberg, N., Kowalczykowski, S.C. en Von Hippel, P.H. (1981) J. Mol. Biol. 145, 105-12
   Howard, F.B. en Miles, H.T. (1982) Biochemistry 21, 6736-6745
   Howard, F.B., Frazier, J. en Miles, H.T. (1977) Biopolymers 16, 791-809

3. Het erratum van T.M. Lohman, verschenen in Biopolymers 24 (1985) blz. 2175, bevat meer onjuistheden dan de oorspronkelijke tekst (Lohman, T.M. (1983) Biopolymers 22, 1697-1713).

4. De methode van Barlos c.s. voor de invoering van tritylgroepen in aminozuren is door een verkeerde opvatting over het reactiemechanisme onnodig gecompliceerd.

   Barlos, K., Papaioannou, D. en Theodoropoulos, D. (1982) J. Org. Chem. 47, 1324-1326
   Kricheldorf, H.R. (1972) Justus Liebigs Ann. Chem. 763, 17-38

5. Uit tijdsafhankelijke NOE-metingen aan oligonucleotiden concluderen Clore en Gronenborn dat de effectieve correlatietijd van de intranucleotide H2'-H2" suiker vector ongeveer een factor drie kleiner is dan die van de intranucleotide H5-H6 base vector en dat interne beweging hiervoor een verklaring kan geven. Er mag echter niet zonder meer van uitgegaan worden dat de kruisrelaxaties tussen het H2' (of H2") proton en overige suikerprotonen te verwaarlozen zijn.

   Clore, G.M. en Gronenborn, A.M. (1984) FEBS Lett. 172, 219-225

6. Goede procesbeheersing in de glasindustrie vergt nauwer begrensde (glas)-specificaties dan zoals die gegeven worden in de eisen van de afnemer.

7. Het gebruik van de "goto"-opdracht in computerprogramma's geschreven in de talen "C" of "Pascal" dient vermeden te worden.

8. Door in Nederland donorcodicils centraal te registreren en deze registratie rechtsgeldigheid te verlenen zal het tekort aan beschikbare organen voor transplantatie-doeleinden verminderen.

H. Bulsink, 20 november 1987.